

SEARCH REQUEST FORM

Requestor's
Name: _____Serial
Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 06-05-03
Searcher: Beverly e4994
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG Suite
_____ STN
_____ Dialog

APS

_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CGN

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:30:47 ; Search time 2758 Seconds
(without alignments)
10562.693 Million cell updates/sec

Title: US-09-482-788-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 caattgatgtctgcagcgc.....cccaagtctctgttacc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|------------|---------------------|
| 1 | 481.6 | 48.1 | 10934 | 8 | FSESYNIA | Z18755 Fusarium sc |
| 2 | 392 | 39.2 | 9633 | 6 | BD013055 | BD013055 Cyclic de |
| 3 | 392 | 39.2 | 9633 | 23 | BD010088 | BD010088 Cyclic de |
| 4 | 360 | 36.0 | 3155 | 8 | COPEPSYNH | X96559 C. oligospor |
| 5 | 339.4 | 33.9 | 46899 | 6 | A36768 | A36768 Sequence 1 |
| 6 | 339.4 | 33.9 | 46899 | 8 | AT050554 | ZR0383 T. niveum (A |
| 7 | 339.4 | 33.9 | 46899 | 6 | TISACTS | 228383 T. niveum (A |
| 8 | 168.6 | 16.8 | 3973 | 6 | A40406 | A40406 Sequence 33 |
| 9 | 106.2 | 10.6 | 40989 | 1 | AF204805 | AF204805 Nostoc sp |
| 10 | 97.6 | 9.8 | 39436 | 1 | ASP269505 | AJ269505 Anabaena |
| 11 | 87.8 | 8.8 | 15582 | 1 | AF455810 | AF455810 Xenorhabd |
| 12 | 86.2 | 8.6 | 3671 | 1 | PSPBSC | X77699 Pseudomonas |
| 13 | 86 | 8.6 | 28587 | 1 | AF047828 | AF047828 Pseudomon |
| 14 | 85.6 | 8.6 | 37856 | 1 | AX024319 | AX024319 Sequence |
| 15 | 85.6 | 8.6 | 37856 | 6 | AX024212 | AX024212 Sequence |
| 16 | 84.8 | 8.5 | 216050 | 1 | AL646076 | AL646076 Ralstonia |
| 17 | 81 | 8.1 | 4771 | 1 | AF021263 | AF021263 Streptomy |
| 18 | 78.4 | 7.8 | 48774 | 1 | AF007865 | AF007865 Bacillus |
| 19 | 77.6 | 7.8 | 7836 | 1 | AF047717 | AF047717 Streptomy |
| 20 | 77.6 | 7.8 | 12744 | 1 | AF204401 | AF204401 Streptomy |
| 21 | 77.4 | 7.7 | 18023 | 1 | AX024392 | AX024392 Sequence |
| 22 | 77.4 | 7.7 | 18023 | 6 | AX024285 | AX024285 Sequence |
| 23 | 77.2 | 7.7 | 3975 | 1 | AMPEPSYNT | X97860 Amycolatops |
| 24 | 75.6 | 7.6 | 22444 | 1 | AE004669 | AE004669 Pseudomon |
| 25 | 75.6 | 7.6 | 31220 | 1 | AE008316 | AE008316 Agrobacte |
| 26 | 75.4 | 7.5 | 7335 | 6 | AX097467 | AX097467 Sequence |
| 27 | 75.4 | 7.5 | 15719 | 1 | AE004667 | AE004667 Pseudomon |
| 28 | 74.8 | 7.5 | 23451 | 1 | AB070952 | AB070952 Streptomy |
| 29 | 74.4 | 7.4 | 10569 | 1 | AE009298 | AE009298 Agrobacte |
| 30 | 74.2 | 7.4 | 77457 | 1 | AF210249 | AF210249 Streptomy |
| 31 | 73.6 | 7.4 | 28798 | 1 | U95370 | U95370 Bacillus li |
| 32 | 73.2 | 7.3 | 28732 | 1 | AF082100 | AF082100 Streptomy |
| 33 | 73 | 7.3 | 28890 | 1 | AF512431 | AF512431 Saccharot |
| 34 | 73 | 7.3 | 37200 | 1 | SC563 | AL035640 Streptomy |
| 35 | 72.4 | 7.2 | 7493 | 1 | PP0310530 | AJ310530 Pseudomon |
| 36 | 72.4 | 7.2 | 32748 | 1 | AB070951 | AB070951 Streptomy |
| 37 | 72 | 7.2 | 13964 | 1 | BLJCHAA | Y10550 B. lichenifo |
| 38 | 72 | 7.2 | 32386 | 1 | BLAJ5061 | AJ005061 Bacillus |
| 39 | 71.8 | 7.2 | 3471 | 6 | AX433752 | AX433752 Sequence |
| 40 | 71.6 | 7.2 | 2792 | 1 | TSF318786 | AJ318786 Tolypothr |
| 41 | 70.4 | 7.0 | 5451 | 6 | AX078430 | AX078430 Sequence |
| 42 | 70.4 | 7.0 | 8992 | 1 | SV117268 | Y17268 Streptomyce |
| 43 | 70 | 7.0 | 7914 | 1 | AF021262 | AF021262 Streptomy |
| 44 | 69.8 | 7.0 | 333500 | 1 | AP003590 | AP003590 Nostoc sp |
| 45 | 69.2 | 6.9 | 230573 | 1 | AF322012S2 | AF322013 Bradyrhiz |

ALIGNMENTS

RESULT 1
FSESYNIA
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

FSESYNIA
Fusarium scirpi esynl gene for enniatin synthetase.
Z18755
Z18755.3 GI:15591913
enniatiin synthetase; esynl gene.
Fusarium equiseti.
Fusarium equiseti
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; mitosporic Hypocreales; Fusarium.
1 (bases 1 to 10934)
Haese, A., Schubert, M., Herrmann, M. and Zocher, R.
Molecular characterization of the enniatin synthetase gene encoding
a multifunctional enzyme catalysing N-methyldeipeptide formation

10934 bp DNA linear PLN 11-SEP-2001

in *Fusarium scirpi*
Mol. Microbiol. 7 (6), 905-914 (1993)
93247491
PUBMED
8483420
2 (bases 1 to 10934)
Haese, A.
Direct Submission
Submitted (24-NOV-1992) Haese A., Technische Universitaet Berlin,
Institut fuer Biochemie u. Mol. Biologie, Franklinstr.29, W-1000
Berlin 10, F.R.G
Revised by [4]
3 (bases 1 to 10934)
Zocher, R.
Direct Submission
Submitted (22-MAR-2000) Zocher R., Technische Universitaet Berlin,
Max- Volmer- Institut fuer Biophysikalische Chemie und Biochemie,
Abteilung Biochemie und Molekulare Biologie Sek. OE2, Franklinstr.
29, D-10587 Berlin, GERMANY
Revised by [5]
4 (bases 1 to 10934)
Zocher, R.
Direct Submission
Submitted (11-SEP-2001) Zocher R., Technische Universitaet Berlin,
Max- Volmer- Institut fuer Biophysikalische Chemie und Biochemie,
Abteilung Biochemie und Molekulare Biologie Sek. OE2, Franklinstr.
29, D-10587 Berlin, GERMANY
On Sep 12, 2001 this sequence version replaced gi:7327885.

FEATURES

source

1. .10934
Location/Qualifiers
/organism="Fusarium equiseti"
/strain="Lambotte et Fautrey"
/isolate="ETH 1536/75"
/db_xref="taxon:61235"
/clone="lambda DASH ESI"
/clone_lib="lambda DASH"
1021. .10416
/genes="esy1"
1021. .10416
/genes="esy1"
/function="synthesizes emniatin"
/codon_start=1
/product="emniatin synthetase"
/protein_id="CAA79245.2"
/db_xref="GI:7327886"
/db_xref="SPTREMBL:Q00869"
/translation="MSLHTPSDQDQPPALASKTLCEQISRALGLGDKKTIENIPPGTFF
QRVDICADDDKQAVGHAVFEIPKIDIDAAALAAAKETVLHTPALRTCTFTSKSDV
LQVLDSEVFWMSGSPVDLKEAVVQDEAAALAGPCNRNRFVLLDPDPTKERLIWT
FSHALVDSTFQRIILRRVLKAYKANDHEHPFETPDSSOATPEEDLPNPKMLKIP
QAADMRAVEFKDHLGSLKCFCLPAFLVSSVYAHPDAAEHRISSSAQOKMSAT
ICRTALALLSRYTHSPEALFGIVTEQTEPLEEQLMDGPTTIVPIRVSCASEQSV
DMSTIDSDQTMROFAHAGLRNIAASAGDESAACGFQTVLLVSDGDAQPASTWEILK
KTEPEPGFIPCTNRALLSCMTSSGAHLTARYDQSIIDAEQMARLLRGLHLQNLQ
TSDLPFVKVDMTQEDWLEIERNWSDSIDAODTLIHSEMLKWTSPNKAAVAAMD
ENTYAEADNVSRQAQIINSIDLGEHAIVPIYFEKSKWVASMLAVLKAGHAFTLID
PDDPPTAQVVOQT SATVALTSKLHRETVQSTVGRCIVVDEFEVKSLPQSSLSASV
KAHDLAYVFTSGTGI PKGIMIEHRSFSSCAIKGPGALGITSDFRALQFGSHATGAC
ILEMTLLHGGCVCIPISDDDNNVLEFINRTNVOLGHTPSYMGTFQFVPEVPGLT
LVLVGEQSSAYNEWAPRVQLNGYQSESSICCVAKISPGSEPNINHAQVGH
WIVDEPNRLAP IGAVGELVIESAGIARDYIVATQDKSPFIKTAPTWNYPKQLPDG
FKIYRTGLDACYASDGSIVCLGRMDSQVIRGORELGAETHLRQMPDDMTIVVEA
VKSDDSTVLTALFAGLGNKSHLDQARTREINAKMEOVLPRESIPAFVSNWL
POTATGKDRRLKRLMGSKILSKQTHSPSQSOQAASISGDTYTKLESWITSIDLE
PGSANNMSTFEMGNSIIAKMWNMARSNLELKVSDIYQNTLAGLKAUIVIGTSLP
LSLIPKVTQRQGVSSQVNAQRMWFLDLQSEGAWSYLPFAVRMRGPVDDVALTRALL
ALEQRHETLRTTFENQDGVQI IHDRLSKELQVLDALDDEGLKTLKYKVETTFDI
TSEAGWSSTLIRLGDHLLISVMHHIISDGSIDVLRRELQVLAALQKDPSSAL
TPLPYSDFAVMQEQAAAEHERQVKKOLADSPAKIPTDFPREDLGSDGAV
VVAIDGELYOKLRCFNKHNSTAFSILLAAERAAHYRLTAVDDAVIGIPIANRWE
LENMIGFFVNTOCRIANVDETDTFESLRVQVRSTTTAAFAHEDVFERVSAIQPGHR
DUSRTPLAQIMEFVHQDLGRLEGEIOSEPIASKAYTRFDVFEHLFQADGLKGS
NFATDLPLFQNVSVFFQLRLGDPQETCISVLPTDGVLEELRLDLEIKRTN
YPRDSVVDVFREQAANPEVIAVTDSSRLTYAELDNKSELLRWLRRRLPETLV

gene

CDS

BASE COUNT 2603 a 3019 c 2726 g 2586 t
ORIGIN
Query Match 48.1%; Score 481.6; DB 8; Length 10934;
Best Local Similarity 68.9%; Pred. No. 1.4e-128;
Matches 693; Conservative 0; Mismatches 304; Indels 9; Gaps 2;
QY 1 CAATTGATGCTTGGCAGCGGATCTCAATCAGCTCTACTCAGCTGGCTCAAGGACTCAA 60
DB 4730 CCATCGACGTTCTCCGCGCGGAACACTCATCCACTTACGCCGCGCTCTGCAGGC 4786
QY 61 AGACCCGGTGTGACGACTCACTCTCTACCTATCCAGTACAGCGACTTTTGCAAAATGGC 120
DB 4787 AGGATCTTCTCCGCACTAAACCCCTACCCATCCAGTACAGCGACTTCCGCGTGTGGC 4846
QY 121 AGA-----AGGACCAATTCAATAGACGAGGAGGAGCACTCACTACTGAAAGAACCAAC 174
DB 4847 AGAAGCAGGAGGCGCCCAAGCAGCTGAGCAGCAGGAGGAGCTCCAGTACTTGGAAAGAACCAAC 4906
QY 175 TCAAGACTCTTCCCGCAGCAAGACTCCGACCCGACTTTTGGCCCGCTGCACTTCTGTCTG 234
DB 4907 TCGCAGATAGTTACCTGCCAAGATCCCTACCGACTTCCCGCTCCAGATCTCTGTCCG 4966
QY 235 GAGACCGAGTGTGCGTACATGTTTACCATCGACGGGAGGTCTTACCAGTCCCTTCCGAGCCT 294
DB 4967 GTGACGACGCGTGTGCGCGTGTGCCATCGACGGGAGGTGTATCAGAACTAAGGGCT 5026
QY 295 TCTGCAACCAACACACACGACTCTTTCGTCGTTCTTCTAGCTGGTTCGCTCCGCGCTC 354
DB 5027 TCTGCAACCAACCAACACGACTCTTTCGTCGTTCTTCCATCTCTGCTGCTTTCGCGCGCGC 5086
QY 355 ATTATCGTCTCAGCTGTGTAAGACGCTGTCTATTGGTACACCAATTTGCGAATTCGCAACC 414
DB 5087 ATTACGCTCTACGCGGTTGACGAGCGCGGTGATCGGCATCCCATTCGCAACCGTAACC 5146
QY 415 GACCTGAATCGGAGGATATCATCGGCTGCTTTGTCAATACGAGTGTATGCGGAATCAACA 474
DB 5147 GCTGGGAGCTGGAGAACATGATTGTTCTTTGTCAACACGCGAGTGTATGCGCATCGCG 5206
QY 475 TAGATCATCAGATACCTTTGGGACTTTCATCAACCAACTCAAGGCTACGACGACGACG 534
DB 5207 TTGACGAGAGGATACATTTGAGAGTCTGGTGGCGGAGGTTCAGATCTACCACTACAGCTG 5266
QY 535 CATTCGAGAGAGGAGGATATTCGGCTTTGAGCGGCTTGTATCAGCTTACAGCTTACAGTCCA 594
DB 5267 CGTTTTCGCGACGAGGATGTCCTTCGAGCGGTGTCGTTTCAGGCGCTTCAGGCTTGGCCATA 5326

| | | | | | | |
|-------------|--|--|---|------|---|--|
| Qy | | 595 | GAGATCTGTCAAGCACACCTCTCGCACAACTCATTTTTCAGTGCATCAACAGAGGACC | 654 | Best Local Similarity 63.3%; Pred. No. 1.6e-102; | |
| Db | | 5327 | GAGATCTCTCGGAACACCGCTGGCACAGATAATGTTTGTCTTCACTCGAGAAGGACC | 5386 | Matches 637; Conservative 0; Mismatches 360; Indels 9; Gaps 2; | |
| Qy | | 655 | TTGGAAGATTCAAGTTCACGGGTCTGAGTCCGCTGACCTGCTGAGCAAGCGTACACTC | 714 | 1 CAATTGATGTCTTGGAGCGCATCTCAATAGCTCTACTCAGCTCGCTCAAGGACTCAA | |
| Db | | 5387 | TTGACGCTTTGAGCTGGAGGTATCCAGCTGAGCTGATCCAGCAAGCGCTACACCA | 5446 | 3710 CTATCGACATCTACGCGGGAACATAAGCAATTTCTATTAGCGCTCTCCGGGGCTC-- | |
| Qy | | 715 | GATTTCATGAGCTTCATCTGTTTCAAGAAACCGACAGCTTAAAGGTAGCTCAACT | 774 | 61 AAGACCCCTGTCTAGCACTCACTCTCTTACCTATCCAGTACAGCGACTTTTGCAAATGGC | |
| Db | | 5447 | GATTTCATGAGCTTCATCTGTTTCAAGAAACCGACAGCTTAAAGGTAGCTCAACT | 5506 | 3768 -TGATCTCTATCGTGTGAGCCCACTCCCACTCCACTACCGGACTTTTCCGTTGGC | |
| Qy | | 775 | TTGCCGATGAGCTTCAAAATGGAGACTGTTGAAAATGTCGTGAGAGTATTTCTTTGAGA | 834 | 121 AGAG-----GACCAATTCATAGAGGAGGAGCAACTCACTTACTTGGGAAGAGAAC | |
| Db | | 5507 | TTGCCACAGATCTCTTCAAGCCGCAAACTATCCAGAAATGTTGTGAGCGTGTTCACAGA | 5566 | 3827 AAAAGCAGTCCGACAGGAGCCGAAACATGAGCGCAACTCGAAATCTGGGTCAAGAGC | |
| Qy | | 835 | TTCTGAGAAAGCGGCTTCAAAGTTCGGGAGACACAGCTCTCAATACTTCTTGTAGTATG | 894 | 175 TCAAGACTCTTCCCGACCAAGATCCCGACGACTTTGCCCGCCCTGCACATCTCTGTG | |
| Db | | 5567 | TTCTACGCCATGGCTTTCACACAGCTGAGAGCTGTATCTCGGTTCTTCCACTGACTGATG | 5626 | 3887 TCGCAGACAGCTGGCGCCCGCAATTTCTTAACCGACTTCCCGGACCCCAACATACTGT | |
| Qy | | 895 | GCATTGTGACTCTTGAATAATTTGATGTTCTCAACGTCACAAATGTCGACTATCCCGAG | 954 | 3947 GTGAAGCAGGTTCCGTCGACGTGACGATCGAAGGCGAACTGTATGAAAGGCTCCAAGA | |
| Db | | 5627 | GAGTCGAGGACTTCGCGAGTTGGATCTGCTGGAATCAAGAGGACTTAACCTCCGCTCGG | 5686 | 295 TCTGCAACGACACAAACAGGACTCTTTCTGCTGTTCTTCTAGCTCGTTCGTTCCGCTC | |
| Qy | | 955 | AATCGAGCTTGGCTGATGTTTCCAGACCCCAAGTCTCTGCTTACCC 1000 | | 4007 TCTGTAAAGTAGAGCAAAATGACGCTTTCCGCGTGTGTTAGGGGCTTTCGCGCGACC | |
| Db | | 5687 | ATTGAGCGTGTAGATGCTCTTCGCGCAACAGCTGCGCGCAACCC 5732 | | 355 ATTATCGTCTCACAGCTGTGAAGAGCTGTGTTGTTGATGTTGATGTTGATGTTGATG | |
| REFERENCE | | BD013055 | 9633 bp DNA linear PAT 02-AUG-2002 | | 4067 ATTATCGTCTCACCGCGCCGGAAGACTCGATCATCGGACGCCCATCGGAACCCGACCC | |
| LOCUS | | BD013055 | Cyclic depsipeptide synthetase and its gene and mass production | | 415 GACCTGAATGGAGGATATCATCGGCTGCTTTGTCAATACGCGAGTGTATCGGAATCAACA | |
| DEFINITION | | BD013055 | system of cyclic depsipeptide. | | 4127 GCCAGGACTTGAARACATGATCGGCTTCTGCTCAACACCCCAATGATGCGAATCACGG | |
| ACCESSION | | BD013055.1 | GI:22093244 | | 475 TAGATCATCAGATACCTTTTGGGACTTTTGAATCAACCAAGTCAAGGCTACGACAGCAG | |
| VERSION | | WO 0118179-A/1. | unidentified. | | 4187 TCGAGCGGACGACACTTTTGAAGCGCTGGTTCGACAAAGTTCGGACCCAGCGGCGGG | |
| KEYWORDS | | unidentified. | unclassified. | | 535 CATTCGAGAAGGAGGATATTCGTTTGAAGCGGTTGATGATGATGATGATGATGATGATG | |
| SOURCE | | unclassified. | unclassified. | | 4247 CATTCGAGCAACAGAGCTGCTTTCGAGCGGCTGCTGAGCGACTCTTCCACGCTCA | |
| ORGANISM | | 1 (bases 1 to 9633) | Mido, N., Okakura, K., Miyamoto, K., Watanabe, M., Yanai, K., Yasutake, T., Aihara, S., Futamura, T., Klein, H., and Murakami, T. | | 595 GAGATCTCTCAAGCACACTCTTCGACAACTCATTTTTCAGTGCATCAACAGAGGACC | |
| REFERENCE | | 1 (bases 1 to 9633) | Mido, N., Okakura, K., Miyamoto, K., Watanabe, M., Yanai, K., Yasutake, T., Aihara, S., Futamura, T., Klein, H., and Murakami, T. | | 4307 GAGACTTATCCGAAACCCACTAGCACAGCTCACCTTCGCTCTTCTTCTCAACAGGACC | |
| AUTHORS | | Yasutake, T., Aihara, S., Futamura, T., Klein, H., and Murakami, T. | Cyclic depsipeptide synthetase and its gene and mass production | | 655 TTGGAAGATTCAAGTTCAGGCTCTCGAGTCTCGAGTCTGCTGCTGCTGCTGCTGCTGCTG | |
| TITLE | | System of cyclic depsipeptide | patent: WO 0118179-A/1 15-MAR-2001; | | 4367 TCGCAAGTTCGAGCTGGAGGCTCTGATGCGGAAACCCGCTCTCGAACAAGGTATACACA | |
| JOURNAL | | MEIJI SEIKA KAISHA LTD. NAOKI MIDO, KAOBU OKAKURA, KOICHI MIYAMOTO, MANABU WATANABE, KOJI YANAI, TETSUYA YASUTAKE, SATO AIHARA, AKAFUMI, FUTAMURA, HORST KLEINKAUF, TAKESHI MURAKAMI | OS Mycelia sterilia | | 715 GATTTGACATGGAGTTCATCTGTTTCAAGAAACCGACAGCTTAAAGGTAGCTCAACT | |
| COMMENT | | PN WO 0118179-A/1 | PD 15-MAR-2001 | | 4427 GGTTCGAGCTGGAGTTTCACTGTTTCCAGAGCCGGAAGACTAAGCGTACGTTGGCAT | |
| | | PF 07-SEP-2000 WO 2000JP006103 | PR 07-SEP-1999 JP 99P 253040.06-APR-2000 JP 00P 104291 PI | | 775 TTCCGATGAGCTGTTTCAAAATGGAGACTGTTTGAATAATGTCGTGAGAGTATTTCTTTGAGA | |
| | | NAOKI MIDO, KAOBU OKAKURA, KOICHI MIYAMOTO, MANABU WATANABE, KOJI YANAI, TETSUYA YASUTAKE, SATO AIHARA, AKAFUMI, FUTAMURA, HORST KLEINKAUF, TAKESHI MURAKAMI | PI TETSUYA YASUTAKE, SATO AIHARA, TAKAFUMI FUTAMURA, HORST KLEINKAUF | | 4487 TTGCGCAGATCTATTCAAGCTGAGACCATAGCAATGTAGTTCGCTATTTTCCAAA | |
| | | PI TAKESHI MURAKAMI | PC C12N9/00, C12N15/52, C12N1/15, C12P21/04 | | 835 TTCTCAGAAACGGGCTTCAAGTTCGGGAGACACAGCTCTCAATACTTCTTTGACTGATG | |
| | | CC peptide synthetase for PF1022 | FH Key Location/Qualifiers | | 4547 TCCTGCGACAAAGGCTATCGCCAGCTTCGAGACTTCCCAATCGCTTCTCCGCTTACCGATG | |
| | | FT CDS (1)..(9633) | FT mat_peptide (13)..(9630). | | 895 GCATTGTGACTCTTGAATAATTTGATGTTTCAAGCTCAACACTGTCGACTATCCCGAG | |
| | | FT Location/Qualifiers | 1..9633 | | 4607 GGTTCAGCGACCTTCGTCGCATGGGCTTCGTTGAGATCGAAGGAGGAGAAATACCCCGGG | |
| | | /organism="unidentified" | /db_xref="taxon:32644" | | 955 AATCGAGCTTGGCTGATGTTTCCAGACCCCAAGTCTCTGCTTACCC 1000 | |
| | | 2318 a 2834 c 2462 g 2019 t | | | 4667 AGTCGAGCGTCTGTCGACGCTTTCGCGCAAGCAGGTGGCGCGCTCACCC 4712 | |
| BASE COUNT | | 2318 a 2834 c 2462 g 2019 t | | | | |
| ORIGIN | | | | | | |
| Query Match | | 39.2%; Score 392; | DB 6; Length 9633; | | | |

[illegible]

| | |
|-----------------------|--|
| RESULT | 5 |
| A36768 | |
| LOCUS | A36768 46899 bp DNA linear PAT 05-MAR-1997 |
| DEFINITION | Sequence 1 from Patent EP0578616. |
| ACCESSION | A36768 |
| VERSION | A36768.1 GI:2294035 |
| KEYWORDS | . |
| SOURCE | Tolypocladium inflatum. |
| ORGANISM | Tolypocladium inflatum Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; |
| REFERENCE | Tolypocladium. 1 (bases 1 to 46899) |
| AUTHORS | Leitner,E., Schneider,E., Schoergendorfer,K. and Weber,G. |
| TITLE | Cytoposrin synthetase |
| JOURNAL | Patent: EP 0578616-A 1 12-JAN-1994; SANDOZ LTD (CH) Other publication JP 625773 940816 Other publication AT 140392 940515 Other publication AT 43793 940415 Other publication AT 398578 941227 Other publication AT 398434 941227. |
| COMMENT | Location/Qualifiers 1..46899 /organism="Tolypocladium inflatum" /strain="ATCC 34921" /db_xref="taxon.29910" |
| FEATURES | BASE COUNT 10651 a 13511 c 12510 g 10227 t ORIGIN |
| Query Match | 33.9%; Score 339.4; DB 6; Length 46899; |
| Best Local Similarity | 60.1%; Pred. No. 3,6e-87; |
| Matches | 603; Conservative 0; Mismatches 391; Indels 9; Gaps 2; |
| Qy | 4 TTGATGTCTTGACGGCGATCTCAATCAGCTACTACTAGCTGCCTCAAGCACTCAAAG 63 |
| Dd | 30226 TTGATCTCTTGACACAGGAGCTCGGTCAATTCTACTCAGCGCTTTACGTGGC---AGGG 30288 |

| FEATURES | source | Location/Qualifiers |
|------------|--------|---|
| CDS | | <p>1..3155</p> <p>/organism="Cylindrotrichum oligospermum"</p> <p>/strain="NRRL 18230"</p> <p>/db_xref="taxon:72418"</p> <p>/clone_lib="PCB11"</p> <p><1..>3155</p> <p>/function="amino acid adenylate forming module"</p> <p>/note="SD2 214-103"</p> <p>/codon_start=2</p> <p>/product="peptolide synthetase"</p> <p>/protein_id="CAA65395.1"</p> <p>/db_xref="GI:1770180"</p> <p>/db_xref="SPTREMBL:P97961"</p> <p>/translation="DPQQQLPVGVNGELVVTGDGLARGYTDVLDQGRFVQIRIKE SVRYATDGRARYRKPDGQIEFGMDQOIKIRGHRIELAEVHAMLHDAVYRQNV IRLQGDPEGRYFQVVRADETVQDDLSRTHGAVNSANWEEQETFEIKTEIKRNLQV LPSVWPARTITLDDMLNANGKVDKDLARRAOTVSKAEKLPASVAPRNEVEVILC EFESVDLGVGEVGVADSFPLDGLSHMATKLAIRSLRNARVSVVEVEDQPVLAIDLAI VTQRGSKPNPILTIAPYSGVPSQAGRLVLDQJNGASWYLPVLAFLVAPRVEGLPHTE ATTAILAEQHEIRLWTFEERDGGVGVVEVHEHTEKLRVIDPADLDGGYLOPLKQ EQTSSFDLASEGRYVSLRIANNHVLISIVMHIIISDGSIVDPLIRRELGLQYLSAALR GYDPOQSPLDGLDFSWQWQKAEQVAEHOLEYWTTIRLADSSPABELLTDLPDPRPT VLISGNAGVQLAIDGSLYBKLRAFCAQYOTTSEAVLLAAFRATHYRLNGAEDATICTPT IANRNPPELENLIGFPVNTQCMRIYVDKDDTETELMQQVRSITTAFAFNQDVPFERIV SALLPESDRISRNPLQVLMFVLHSDQGMQIEGVVGPVPTAASRPLDLEFLHFQK VRELGGTFLVATDLPEDPTIRGMVTFQVFLHVLDDQHPPTITSLPANGLAGLNCNM LIDIERSDYPRESSLVDVFRQGVAACPDATAVIDSSRLTYTQDQRSDQSVAGWLRRR HMAAETLVGLVAPRSCOTTVAFLGILKANLAVLPIDVNPAAIRIEAILSAVBGHLKVL LGSDDAAPKRVQDVMVMDVDTGLCYDILDDYANSVARESPSTSLAVVMTSGSTGCKPK GYNIBHRAVRLVLKNSVNSIILPMTFPRVAHLNGLFDSVQEVYATALLNGRLVLCID FTTLDNSALETFVMEQEVQRVAMLTPTALLKLCIADIPSVLSGVDLVFNAGDRFDSRDAL AAQAALIRCCVYNAWGPTENG1"</p> |
| BASE COUNT | 683 a | 914 c 859 g 699 t |

| Query Match | 36.0%; | Score 360; | DB 8; | Length 3155; |
|-----------------------|-----------------|--|-----------|--------------|
| Best Local Similarity | 61.3%; | Pred. No. 2.9e-93; | | |
| Matches 617; | Conservative 0; | Mismatches 380; | Indels 9; | Gaps 2; |
| QY | 1 | CAATTGATGCTTTGGCAGCGGATCTCAATCAGCTCTACTCAGCTGGCTCAAGGACTCAA | 60 | |
| Db | | | | |
| QY | 1299 | CTATCGATCTCATACCCGAGAACTAGGACAACATAATCCGAGCGCTTCGTGGCT--- | 1355 | |
| Db | | | | |
| QY | 61 | AAGACCGCGTGTGAGCACTCACTCTCTACCTATCCAGTACAGCGACTTTTGGAAAATGCG | 120 | |
| Db | | | | |
| QY | 1356 | AGACCCCTCAATCACAGCTTAGCCCGCTTCCCTATCCAGTACCGGACTTTTCAGTGTGGC | 1415 | |
| Db | | | | |
| QY | 121 | AGAA-----GACCAATTCATAGAGGAGGAGCAACTCAACTACTGGAAGAACAAC | 174 | |
| Db | | | | |
| QY | 1416 | AAGAGCGGCGGAGCAGGTGGCGGAGCATGAGCGCAGCTCGAGTACTGGACAACACGAT | 1475 | |
| Db | | | | |
| QY | 175 | TCAAAAGACTTTCCCGCAGCAAGATCCCGACCGACTTTGCCCGCCCTGCACITCTGCTCG | 234 | |
| Db | | | | |
| QY | 1476 | TGGCCGACGACTCTCCCGTGAGCTTCTGACTGACCTTCCAGACCGACCGTGCTATCTG | 1535 | |
| Db | | | | |
| QY | 235 | GAGACGCAAGTTCGGTACATGTATACCATCGACGGCGAGCTCTACCAAGTCCCTTCGAGCCT | 294 | |
| Db | | | | |
| QY | 1536 | GCAATCGCGGCTCGTCCAGCTCGCCATTGACGGTTCGTTGTACGAGAAGCTCCGGGCAT | 1595 | |
| Db | | | | |
| QY | 295 | TCTGCAACGAACAACAACAGCACTCTTTCTCGTCTTCTTAGCTGGCTTCGCTGCCGCTC | 354 | |
| Db | | | | |
| QY | 1596 | TCTCGCGAGCCTACCAGAGACACATGCTTCGCGTGTGCTCGCGCATTCGCGCCACTC | 1655 | |
| Db | | | | |
| QY | 355 | ATTATCGTCTCACAGCTGTTGAAGACGCTGTCAATTCGGTACACCAATTTGGAAATCGCAACG | 414 | |
| Db | | | | |
| QY | 1656 | ACTACGCGCTTCACAGCGCGGAGAGCGCCACCATTTGGTACGCCCATTTGCCAACCGGAATC | 1715 | |
| Db | | | | |
| QY | 415 | GACCTCAACTGGAGGATATCATCGGCTGCTTTTGCTCAATACGCAAGTCTATGCGGAATCAACA | 474 | |
| Db | | | | |
| QY | 1716 | GGCCGGAGCTGGAAATCTGATCGGCTCTTCTGCTCAACACCCCAATGCATGCGCATCACCG | 1775 | |
| Db | | | | |
| QY | 475 | TAGATCATCACGATACCTTTGGGACTTTGATCAACCAAGTCAAGGCTACGAGACACAGAG | 534 | |
| Db | | | | |

DAGRRVSVNARHOSQHGALDAVPHHLKPAEDSRVLKPTDHOGRPLKSLNQPILL
PAGRRRELLIREGLQTLPLPIPIQSITLIDRMLNANGKVDRELRARAKITQKSK
PVEDIVPRNSVETGCKGTDLVGEVITIDNFFNLGHSLSMATKLAARLGRQLNTR
ISVRDVPQVADLAIVIQNSAPHEPIKPADYTPVQPSFAQGRVFLDQLNGLVTR
WYMLPLGIRLGRSLRVDALATASALRQREPLRTTTEEDGVCVQVODHRPKDLRI
IDLSTOPKDAYLAVLKEHOTLPLDPEGRVVALIRLGEHHILSLVMHHILSDGHS
VEVLFDEPHRFYSALROODPMQOILLPTQYRDFPAWQKTEEOVAHQRLDITWTH
LASTPAELLTDLPRLSILGRANELPLTIANRPELENMIFGVNTQCMRIAEENDNFESLVRVRS
AAHRTUGADEATLGTPIANRPELENMIFGVNTQCMRIAEENDNFESLVRVRS
TATSAFANOVPESIVSSILPGSRDASRNPLQVILAHVSQDGLKLTLEGLRDEAV
DSAISVDFEHLFEHADRSLGSVLYAKLEKLTIESVSVFLETLRALDOPILTP
LAVPLTDGVEVETASKGLDPRDYPRDANIVVEQOHRVATPDATAVKDATSILTY
AQLDQSDRLAIHLRSHRMPETLVGLVAPRSCETIIFAMGIMKANLAIPLDINSPA
ARLSILSVAENKLVGLVGTAPQENVEVEAVGLOEILAGTGLDKTQSNARPSA
TSLAYLFTSGSTKPKGVNHRSTRAPKSNVSKLPGQARVAHLANIAFASIW
EIAITLLNGATLVGLDYHLDCTRLKEVERESITVITLMPALLKQCAVEIPELIAL
LDLITLGGDVRGCHDAMRARSILVKIMFGSGYPTENTVISTIVEADAMFVNGVPIG
KTVNSGAYVMDRNOOLVPSGVVGLVITDGLARGVTDPSLANKRFLYITVNGESIR
AYTGDRVRYRPHDLOIEFFGRMDQVVKIRGHRIEPEGEVESALLSHNSVQDAVICA
PADQSGAEMVAFVAKNTEDEDTQEEAAVDQVQGWETHFETAAYSEVKDIDROSEVN
DFMGWTSMDGSEIDKTDHMLNDTMRMLTDAREPCHVLEIGTGTGMWFMFLAKCPG
LOQYVGFEPSSAAQFPAALKDGSRIVHVGTATDINKAGTIOPLRVILNSV
AOYFPTPEYLRANEEELLDPAFETSLRTRLEKIKHVELLPKTKMATNLSKYRYAAVLH
ROMIYELEANEELLDPAFETSLRTRLEKIKHVELLPKTKMATNLSKYRYAAVLH
VRGSRQSTIHOVSPNADWDAADGLDQRLINLLEKHDAGTVAIGNIPYSKITIVER
FVNSSEDDMEGQNSLGSAAVAAVMAAQCPSLDAMDVKELIAOEAGYQVEVSA
ROMSNGALDAIEPHFEPPEKAGTLEFPTDYEGRNVTNLRNLSIQSRRLGTQI
REKLTQLLPPYMPISRIWLDQPMVNNNGKIDRKLKRAIVAPKPSAATRVAPRNE
IEALRDEDFEGLTVSVLDNDFDLGHSLSMATKLAARVSRRLDASHIKDVFDPV
LADLAALSIORESAPHEPIPORPYTPGAQSFAGRLWFLDOLNGLATWYMLPLAIR
GOLVAALSALPALRRHETLRTTEESDGVQVIVGEARNSDLRVHDSVGTGGDF
LEVLRQETVPLDSESGRNVCLVXTGEEDHLSIVMHHIITYDGSVLDLRLGELGOF
YSAALRGQDPLHANPLPTQYRDFPAWQKTEEOVAHQRLDITWTHVSTPAELLT
DLRPRSLISGRAGSVITQESVYALQSFCTRSVTTFVLLTFRHIAFRHTAVDD
ATICTPIANRNPETLVGCFVNTQCMRISIAADDNFEGLROVRNATAAAYANDV
PFRRIYALVPSGRNTSRNPLVOLMEAVOSVEDYDOVRLEGLSVNMGCEASPRFME
NLESLLLEMTSDYPRDRTVDFRQEAICPDSTAVKSDSOLTAQDEQSDRA
AWHERHMAESLUSVLSRSCETIIFAMGIMKANLAIPLDVAIPADRLAAILDVE
GERLLGAGVPOPGTOIPRLSTAYIAEALSALTVDVTSIQPSATSILAYVIFTSGS
TGKPGKMIHEHRIQVLRDNTVNVFPESSGALPVSHFNLAAMDATWELIYAVLNG
LVCIDRDMTDLAALNSTFRKENVRAAFETPAFLKQCLAETPELVANLEIHTAGDR
LDPGDANLAGTKAGGIFNVGLHTENTAVSTFVPVGEETVNGVPGSINSXAYI
IDRHOXLPAGVNGELILITGDVGARGYTSALNKDRFVIDINKSTWSRTSKARY
PRDQGLEFFGRMDQVVKIRGHRIEPEGEVELTLLDHKSALAAVTVVRPNPDPENIA
FITIDAEEDVTHKALYKHLQGLTPAYMIPSHLVIDLQMPVTDNGVDRKDLALRQA
VOKRRSTVARDVDEAVLCEEYSNLLEVEVITDGFDFLGHSHLLATKLAARLSR
QNLNRSVARDVDEAVLCEEYSNLLEVEVITDGFDFLGHSHLLATKLAARLSR
NLGASWLMPEALMRGEPLOTKALAVANALVHRHEALRTTEDHDSGVGVVOTOPSS
QDLRIITLSDAVDDTAYLAALRQETTAEDLSEPGRVSLRLRIGDDDTILSLVMHHI
ISDGTVDVRLQELGQFYSAALIRGQEPLESLPIQYRDFPAWQKTEEOVAHQRLDITWTH
KYWSQOLADSTCEFTLDPRLSPILSGEADAPMVIDGTVYQLLTDPCRTHOVTSFV
LLAAAFTHRYRTGLTDATVGTPIANRNPETLVGCFVNTQCMRIAEENDNFESLVRVRS
VOOVRTLTTEANQDVPPEQVITLPGSRDTSRNPLQVMAFLOQSDOLGRIQLEG
MDEALTEPLSLRDLVLEHFOVGLKSLSLYSTDLEFVETIRGLVDVLEITRGL
EOPKORLMAMPITDGTITKLRODGLTLVAPYAPRESSVIDLFRQVAAAPDAIAMD
SSTLYADLDGSGNKLAMHICQNMMAPELTVAPRSCETIIFAMGIMKANLAIPLDINSPA
DVNAPARLAISSVAGHKLVLQVAGHKLTMADTELVOIDEALASSSSDHEQI
HAGSPDTSLAYMFTSGSTKPKGVNHRSTRAPKSNVSKLPGQARVAHLANIAFASIW
AFDISVQETIYALLNGTGLVCLDLYTLDSKILYNVFEAVQVNAAMFTVLLKQCLGN
MPALISRLSVLFNVGDRDLADHDAVASGLIQDAVYNAYGPTENGWQSTWYKDVNPEP
VNGVPIGRSITNSGAYVMDRNOOLVPSGVVGLVITDGLARGVTDPSLANKRFLYITVNGESIR
IGEBENIKAYTGDVRYRPHDLOIEFFGRMDQVVKIRGHRIEPEGEVESALLSHNSVQDAVICA
DAAVLURKPANPEMIAFITSOEDETIEQHSNKOVOGGEHFDVRYADIKDLDT
TFCHDPLTFMIDYDIPVNMKEWDELDTASLLDNRPPHGLEIGAGTGMLSNLG
KYVDGLQYKVLGPSPAAIFVNAEVSPLSLAGKARVLVGTALDGLDKNEIOPELV
VINSVAQYFPTSEYLIKVVNAVEVPSVKRPFEDIRSOALNRDLFLAARVAILGDNA
SKEQIREKTAELAESEEEELLVDPAFVLSRSLQPNIKHVEVLPKLMKATNELSSSTRYA
AVLHSHNEEQILLQDIDPTAWDAVFAATQKDSQGLNRLNLLQGRDDVMIAVGNIPYSK

Query Match

Best Local Similarity 33.9%; Score 339.4; DB 8; Length 46899;

Matches 603; Conservative 0; Mismatches 391; Indels 9; Gaps 2;

RESULT 8

A40406/c

LOCUS

| | | | |
|----|-------|--|-------|
| QY | 4 | TTGATGCTTTCGACGGGATCTCAATCAGCTTCTACTCAGCTGCGCTCAAGGACTCAAAAG | 63 |
| Db | 30226 | TTGATATCCTCGACAGAGGCTCGTCAATCTACTCAGCGCTTACGTGGC---AGGG | 30282 |
| QY | 64 | ACCGGCTGTGAGCACTCACTCTCTACTATCCAGTACAGGACTTTCGAAAATGGCAGA | 123 |
| Db | 30283 | ATCCGTTATCTCAGGTCGAAGCCCTCCCAATAATACTGCTGACTTTTCGGCTTGGCAGA | 30342 |
| QY | 124 | AGGA-----CCAAATTCATAGACAGAGAGCAACTCAACTACTGGAAGAGCAACTCA | 177 |
| Db | 30343 | AGGAAGCTGCCCAAGTTGCCGAGCATGAGAGGAGCTCGGCTACTGGGAGAACCAAGTTAG | 30402 |
| QY | 178 | AAGACTCTTCCCGACAGAAAGATCCGACCGACTTTGCCCGCCCTGCACTTCTGTCTGGAG | 237 |
| Db | 30403 | CTGACAGTACTCCCGGTGAGCTTCTGACCGACTTTCCCGCCGACAGTCTCTGAGTGGGA | 30462 |
| QY | 238 | ACGAGGTTCCGTACATGTTACCATCAGCGGAGCTCTACAGTCCCTTCGAGCCCTCT | 297 |
| Db | 30463 | AGGCTGGTGTCTATCCCGGTCACTTTCGCTGCTTCTTCTAGCTGGTTCGGTCCGCTCAAT | 30522 |
| QY | 298 | GCAAGCAACAACAACAGCACTCTTTCGCTGCTTCTTCTAGCTGGTTCGGTCCGCTCAAT | 357 |
| Db | 30523 | CAAAGGAGCCCAAGTAACTCTGTTCTCGGTGCTATTAAACAGCGCTTCCGGGCCACACT | 30582 |
| QY | 358 | ATCGTCTCACAGCTGTTTGAAGAGCGTCTATTGGTACACCAATTCGGAATCGCAACCGAC | 417 |
| Db | 30583 | TTGCTCTACTGTTGACAGAGATGCTACGATCGTACCCCAATTCGAAATCGCAACCGCG | 30642 |
| QY | 418 | CTGAAGTGGAGATATCATCGGCTGCTTTCGTAATAGCGAGTGTATCGGAATCAACATAG | 477 |
| Db | 30643 | CAGAATCTCGAGCATCAATTCGATTCCTTCAACACCCCAATTCGATCGCTTCTCTCTCG | 30702 |
| QY | 478 | ATCATCAGGATACCTTTGGGACTTTGATCAACCAAGTCAAGGCTACGACACGACAGCAT | 537 |
| Db | 30703 | ATACCGGACGACATTCGAATCCCTAGTCAGCATGTTTCGGTCCGTGGCTACAGATCGCT | 30762 |
| QY | 538 | TCGAGAACGAGGATATTCGCTTTGAGCGCTTGTATCAGCACTACAGCCCTGGATCCAG | 597 |
| Db | 30763 | ATTCCAATCAGGATATTCCTTCCGAAGGATCGTTCGGCACTTCTCCCTGGCTCGAGAG | 30822 |
| QY | 598 | ATCTGTCAAGCACACCTCTCGCAACTCATTTTTCGAGTGCACACAGAGACCTTG | 657 |
| Db | 30823 | ATGCCTCAGGAGCCCACTAATCCAGCTTATGTTTGCCTTGCACCTCACAGCAGATCTCG | 30882 |
| QY | 658 | GAAGATTCAAGTTCCAGGCTCTCGAGTCCGTACTCTGTGCTAGCAAAAGCGTACACTCGAT | 717 |
| Db | 30883 | GGAACTTACTCTCGAAGGACTCGAGTACGAGCGCTCCCAACAGCGTCCGACACGCTT | 30942 |
| QY | 718 | TTGACATGAGTTCCTATCTGTTTCAAGAACCCGACAGCCCTTAAAGGTAGCGTCAACTTTG | 777 |
| Db | 30943 | TCGACATGAGTTCCTACCTGTTTCCAAAGGCTTAAAGCTGAGTGGTTCATTAATCTCTTG | 31002 |
| QY | 778 | CCGATGAGCTGTCAAAATGGAGACTTGTGAAATGTCTGTCAGAGTATTTCTTTCAGATTC | 837 |
| Db | 31003 | CCGATGAGCTTCTTTCAGCTGTAAACCAATCAACAGCTCTGCTGACTGTGTTTCCAGGAGATAC | 31062 |
| QY | 838 | TGAGAAACGGGCTTCAAAAGTTCCGCGACACCACTCAATCTTCTTTCGACTGATGGCA | 897 |
| Db | 31063 | TCGAGCGCGCTTCGACCAACCCCAAGTCTCCATTTCTACTATGCCCCCTGACTGATGGGT | 31122 |
| QY | 898 | TTGTGACTTTGAAAATTTGGATGTTCTCAACGTCAAAACATGTCTGACTATCCCCGAGAAT | 957 |
| Db | 31123 | TGATGATCTCGAGAACTGGGCTTCTGTAATCGAGAGCAGCAACTTCCCTCCGCGACT | 31182 |
| QY | 958 | CGAGCTTGCTGATGCTTCTCCAGACCCCAAGTCTCTGCTTACCC | 1000 |
| Db | 31183 | ACTCGGTTGTCAGCTCTTTCGAGACAGGAGGCTGCTGCCAATCC | 31225 |

A40406 3973 bp DNA linear PAT 05-MAR-1997

DEFINITION Sequence 33 from Patent WO9425606.
ACCESSION A40406
VERSION A40406.1 GI:2296446
KEYWORDS Tolypocladium inflatum.
SOURCE Tolypocladium inflatum.
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Tolypocladium.
REFERENCE 1 (bases 1 to 3973)
AUTHORS Koehler, H.P., Schneider-Scherzer, E., Schoergendorfer, K. and Weber, G.
TITLE RECOMBINANT ALANINE RACEMASE AND GAPDH FROM TOLYPOCLADIUM
JOURNAL Patent: WO 9425606-A 33 10-NOV-1994;
SANDOZ AG (AT)
FEATURES
 Location/Qualifiers
 1. 3973
 /organism="Tolypocladium inflatum"
 /strain="ATCC 34921"
 /db_xref="taxon:29910"
 join(1953..2034,2111..3180)
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAA02484.1"
 /db_xref="GI:2296447"
 translation="MKICPILNSKAGHWGDPDLTGSAPDFRSDAITTPSHGMLAIAAG
 ATLNDVYGEDQLTRSFEEHMASICGEALFVSGTMAQIALGALSRRCTGYLADA
 TSHVHFPEAGLAGLSGAIQVVRPANGHYLTDDVERHAVTTNNIERLPTSIISLEN
 TAGHSLPLOBELRWKAWDEQETAIHIDGARIMHAGVGGSLKEIETAGCCDAMTISF
 KGKLAAPIGSAIVGSRDLLIARLRQSIGGVRRKAGAIYAAQAMIENTFGPDVDIT
 RGIORTHLVARVANWMTSGGLLRPMVNLVWLDUKSAGLDKSTLDTAAQHRDIR
 IRAPRIVLHOISTDAMRRLVAFVEVLEKKASQSLRSRAERAGLERL"
BASE COUNT 917 a 1036 c 1103 g 917 t
ORIGIN
Query Match 16.8%; Score 168.6; DB 6; Length 3973;
Best Local Similarity 56.4%; Pred. No. 1.4e-37;
Matches 337; Conservative 0; Mismatches 254; Indels 6; Gaps 1;
QY 4 TTGATGCTTGGCAGCGCATCTCAATCAGCTCTACTCAGCTCGCTCAAGGACTCAAAAG 63
Db TCGACATTTATCAGCAGGAGCTTGAGAACTCTACACGGCGCCTCGCAGGGGAATCGA 539
QY 64 ACCGCTCTCAGCACTACTCTCTACCTATCCAGTACAGGCACTTTGCAAAATGGCAGA 123
Db TTTTCGGCTGTCCCTTGGGTCCAAATCCCATCTCAATACCGTCACTGACGACTTGGCAGA 479
QY 124 -----AGGACCAATTATAGCAGGAGAGAGCACTCACTACTGGAAGCAACTCA 177
Db ACCGAGCAGCAGGCTCGCTGAGCAGGAGGAGCAGCTCGGATACCTGAGATCGAGAGCTCG 419
QY 178 AAGACTCTTCCCGCAGCAAGATCCCGACCGGACTTTGCCCGCGCTCTGCTCTGGAG 237
Db ATACAACACACCGCGCGAGCTCTCTCAGAGCTTCCCGCGCCACCTATCCATCTGGCG 359
QY 238 AGCAGGTTGGTACATGTTACCTACGAGGGGAGCTCTACAGTCCCTTCGAGCCTTCT 297
Db TTTTCGGCTGTCCCTTGGGTCCAAATCCCATCTCAATACCGTCACTGACGACTTGGCAGA 479
QY 298 GCAACGACACACACGACCTTTTCGTCGTTCTTCTAGCTCGCTCGCTGCCGCTCAT 357
Db GCGGCTCCGACAGTAACCGCTTACGCGGCTGCTGCTGCGAGCGTTTCGCGTGGCGACT 239
QY 358 ATCGCTCACAGCTGTTGAAGACGCTGTCAATGTTGATACCAATTCGGAATCGCAACCGAC 417
Db TTTTCTTCTGAGGATATATCGGCTGCTTTGTAATACGAGCTGTATGCAATCAACATAG 477
QY 418 CTGAACCTGAGGATATATCGGCTGCTTTGTAATACGAGCTGTATGCAATCAACATAG 477
Db TTTTCTTCTGAGGATATATCGGCTGCTTTGTAATACGAGCTGTATGCAATCAACATAG 477
QY 178 CGGAGCTGAGAAACATGCTGGCTCCCTTGGCCACTCTCGAGTGCATGCGAGTGGCTCG 119
Db ATCATCAGATACCTTTGGGACTTTGATCAACCAAGTCAAGCTACGAGCAGCAGGAT 537
QY 118 ACGAGGACGACACCTTTCGAGTGGTGGCTGCGGCGAGATCATGTCCGCTATGACAGAGC 59

QY 538 TCGAGAACGAGGATATCCGTTTGAGCGCGTGTATCAGCACTCAGCGCTGGATCCA 594
Db 58 ATGCCAACCGGAGCGTCCCTTTGAGCGCATCGTCTCGCTGCTGCCCGGCGTGA 2
RESULT 9
LOCUS AF204805 40989 bp DNA linear BCT 15-DEC-1999
DEFINITION Nostoc sp. GSV224 nostopeptolide biosynthetic gene cluster,
complete sequence.
ACCESSION AF204805
VERSION AF204805.2 GI:6563396
KEYWORDS Nostoc sp. GSV224.
SOURCE Nostoc sp. GSV224.
ORGANISM Nostoc sp. GSV224
REFERENCE 1 (bases 1 to 40989)
AUTHORS Hoffmann, D., Hevel, J.M. and Moore, R.E.
TITLE Characterization of the nostopeptolide biosynthetic gene cluster of
Nostoc sp. GSV224
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 16388)
AUTHORS Hoffmann, D., Hevel, J.M. and Moore, R.E.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1999) Chemistry, University of Hawaii at Manoa,
2545 McCarthy Mall, Honolulu, HI 96822, USA
REFERENCE 3 (bases 1 to 40989)
AUTHORS Hoffmann, D., Hevel, J.M. and Moore, R.E.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1999) Chemistry, University of Hawaii at Manoa,
2545 McCarthy Mall, Honolulu, HI 96822, USA
REMARK Sequence update by submitter
FEATURES
 Location/Qualifiers
 1. 40989
 /organism="Nostoc sp. GSV224"
 /strain="GSV224"
 /db_xref="taxon:76334"
 1370..14509
 /feature="nosa"
 1370..14509
 /feature="nosa"
 /function="activation of L-isoleucine, L-serine, (2S,
 4S)-4-methylproline and leucine"
 /note="peptide synthetase; contains 4 modules"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="Nosa"
 /db_xref="GI:6563397"
 /translation="MNVIKINPLSYGOEAMFVSQIATGIAVNIETAKINCLN
 IAAVNRWEKIEHPILRTYTHIEGRPVQVQNOHKFNLVANASDSEKOLKEKI
 FAITDRPENLEKSDILNVLFTFRSKHEHLLITLMHHIAGDMWSFDLLSEFOTLTAKE
 IEQVSEQIOTAPDSLTENKSYLDFRQWSEVLSARGEKLMQYQKLAGELPTLN
 PADKPRPVQTYGTGYSGLKDLERLIKVLALASGTSLYQVLLAFAVLLIYRTNQ
 TDILGSPMRGRWGREFKEIVGVNANLVRSVOENATFEFLAQVSKTVREAQHQ
 DYPFSLAEQLQPDRISSPLCQVSTFQWHRWCEPTNSLHSGEOLLEIKPILLG
 HQRGADFNLNLMWEGAVQLQWQITDLPASTITRMAGHFVTLLEGVVNPQEQI
 WQLPLTTEQQLLVQNDVGLQNKCTHQLFEQVDSFNPVAVVYENQQLTYQ
 QLNVRANGLALQSLGVENLVGLCVRSLEVMVGLGLKAGGAVPLDPEYPAE
 RLSFMULTHVKLVLTQKLVESLPHQARVCLDTDWHISIQANODNLNSTVAENL
 SVYVTSGSTGTGKVAVTHQAVNRLVNTNYIQLTADCEVAQANAIADFAATFRWG
 ALLSGAKLVITTKSVLLPEEPVNIHKHKISVLELTALFNQLASVVPQAFSSRLYL
 LFGGKAVDPKWLVDKGAQPHLLHYGPTENTTFSSVLVLEDTATTATPIGRAI
 ANTQIYILDQSLQPVGPGEIYLGAGLAGQYENRPELTKEFIPNPFKAGRSRQ
 EAGGRQKVEGSKLYKTGLARYLADGNIEYLRIDNQVKIRGFIELSEIARVLSQ
 GDVQVSVIVREDPTGDKRLVAVVHONCKPTISELROFLKAKLPDYMVPAIVILE
 SPLTPNCKVDRLALPTDLHSEQODKFWAPNPIEEMALLLWQVLVLELLGTHDNF
 FELGHSLATQVLSRIRNRFKVELPREFARVAELAQIRLQOQDLESEAPI
 LPKVRDAELPLSYAQORLWFLDQLEPNSAFNIPMALRVLGNLPALQEQSLLEIHR
 HEALRTNFTVDGKPSQIOTRREGQGTNRQGTVSIVDFKDLSTNEQEIASKLARQ

PRYAGVNTSWALNNOEKTGCTVTHHMAAAMDAGDILKQVILIDIADETTALTALNGKCY
ESAINAFAPVDELSSGTFVATVNLNMQRTYVFRKRLRAGGIISWKRCAITVETDALIR
ALDFGSPNLPGRKLAIDSNFVLSKVEYGNLSNPYPGTITNIEYTYITVSTAYD
IALRVLAINQSALSYLSYVEKFGQVYCDLEPNQVQKIEKFDQSIKHEAFWE
RLGLTLESITPEAKQTSALHLKEPOVARARMPVDEAITLWLRHPQWHRSDFLAAAF
ITYLARIGSGGCDIGDKLELOROLVGLSFASVYVYNIDYQSFPAALKKQREF
TOLPLAVRVDVTRYPSLRSLSHRGSEQFPVYVVERVEKLEIDYQSPSLDFTLIS
DGKKCWFTNDVINDSIRAMQEQFTVFLQGLTEPDQCIAYLPLSQORREILLE
WNTDQVDDPKDCIHOLESQVOLTDAVAVFVFNQOYLHELNCRAQLAHYLSQIG
VGTDLIVCERSLEWVGLGLIKAGGAYVLPDEYPTERLTFMLADLAQVSLVLTQ
QHLVEKLPNOEVPVCLTDMLVICSSQSPITEVQPGNLAVIYVTSGSTGPKGVM
LSHNLNCNPMWQATFPLETKDKVQKTPFGDASWREYFAPLLAGGOLLIAPKGGH
TDSAYLLRLAACQVITVOLVPSLLOMLLEOGGIECHSKLHVFCGGEVLPVALLBGL
LSKLDVNLHNLGTECTDATPCNOREIAQIVPGRISNTQYILDNQLOALPV
GVPELHISAGARGYLNRPTEQKFTIANPSTPGSRLYKTDGLARYLPNGNIEY
LGRIDNQVIRKTRFELAEAEVNLNQDDVQVACIAREDDPGNKRLVGIYVPSQMT
CTVSELPEQLTKPLDYVFNFIWMLDSLPLTPNGIKRGALPVPDLQOGOTKYVAP
RTPEETLSLMAQLVLEQVITHDNFTFGHSLATOLVSRITSFKVELPLRSIF
MTLLAADFLLTYTGTEIDLVGTPIANRDRSEIEGLIGFVNTLVNRSNLSNPFA
ELITRVRSMAISAYAHONPLEMVEALOPERDLSHTPLFOVMALONPXLSEVELSG
LSLSSLPLEGAKFDLTAMENTGNGLGVWEYNTDLDLFIESTIEMTGHFLTLSSY
VANPLQISLPULTTEVERKQOLLIENHTQVNPVDOCIHQLESFQVTPDAVAVF
EEQLYTHELNCRAQLAHYLSQIGADVLGVICVERSWEMIVGLLTIKAGGAYVP
LDPEYPPERLFTILTQVKVLLTOQOLVNLKPAHTAQVCLDTDEKTIQNSNPV
NTATSNLAVIYVTSGSTGQPKVNLVNHHTVRLFAATNSWYKFNQSDQVMTFHSVAF
DFSWEIAGVGLVGLVYVPLTRSPHSFYKLLSOEQVYTLNQTSPAFROLIIOAO
SIGMSNLRLVIFEGEALLESLOPWFERRHQDKQLVNMVGIETTVVHYTRPUSL
DDNRASVIGRIPDLQVYVLDQHLQVPVIGVPGEMYVGGEGVTRGYLRHDLTQOR
FISNFORSGGRLRYTGDLARYLPNGELEYLRIDQVQKIRGFIEIGEIEGLLAQ
HPAVESVYVREDESNKRLVAYVYVVOAQAAPTAEELSKKPLKPLDYMPIAVIL
ESLPLSNKIDRSLPVPSRAGIEESLVAAPTVEEKLAAQIWAQVLRVQVGIHDL
FFELGGHSLATOLVSRITSFKVELPLRSFPAVLAELIQLQLOQNIKIPVP
ILPRKKNKS"
16122..31304
/gene="adpB"
16122..31304
/gene="adpB"
/codon_start=1
/transl_table=11
/product="peptide synthetase"
/protein_id="CAC01604.1"
/db_xref="GI:9715734"
/translation="MQNHNEPLSYAQORLWFLDQLOPNSPLYNIPLALRLVGLNLR
AALQSKETIDRHEALRTNMTIDGQPAQIITQTEINMTVSVVDLQHLSTLVLSVEV
TEQESASQOLAQKAOIQPDLASESLVRLATLVLSSETHLLVCHHIIVSDGWSGVF
AHELALYNAVYTGQRTPLEPLIQYADFAINQRELQGLDLSQLSYQQQKEAPA
LUSLPTDRPRPAPVTVGAHQKFAVSVELTKLKSQOEVTLEMTLLAAFDLILYR
YTGSDIVYGAPITANNRYGEIEALIGFFVNTLVMTNLEGNRFSSELLTRVRLMADA
YHQDLPLFMEALQERDLSYVPLFVFGVGLQNPNSQVELTGLTVPLVSVYTA
KCDLTAMENTGTLGIVWEYNTDFTGCTIERMTHGFVTLLEGIVANPLRSOLPL
LVEVEQQLLVEMNDQVYDPODKCIHOLEFEQCLRTPDVAVFVFNQOYLHELNCR
ANQALHLSHGADVGLVIGICVERSELMIVGLLAIKAGGAYVLPDPYPPQRQFM
LKDGSYLVLTQSLFDLQPIAEQNPQYVICLDQESFSELTNDPNPOSTNNLAVY
VSGACLYNIETLPSQLVDFLADHKISHFSFLSPSALSVPQATLPDLOVLSVATY
ASSAELVWGTGRFRFNAYGPTESTVAATIAICOPNGRKPPIGRPIANQIYILDEY
LOLVPVGVPEGLHIGAGLARGYLRNPELTQKFTPNPEDNSKFKIPYRNADQVSK
LYKTDGLARYLPDGNIEYLRIDNQVKIRGFIEIGEIEAVLSQLDQVQASCVIRED
ITGDKRLVAYIVPQREVPTVTELQKFLAKLPEYMPNPAFVILLESPLTPNGKIDHR
ALPTDPLHSELKHKVAPRTPVEILAQIWAQVMPVQVQIEDNFTSGHSLATQL
VSHIRNIFVELPLRELFAPTAIELMPLIHLQOQDLELTPPLIPRAENAEILLISF
AQNRWFLDQEPNSALYNIPALRLVGLNQAALQESLOEIIHRHEALRTNLITVDG
QPROIOTATNTLTVAVDLKSLSTAQETAAQOLVQOQALQPDLEKQALVRLVVL
SYTEHLVCHHIIVSDGWSGVYVQELATLYNAVIGOPANLAPLIQYADLQWR
QMLAGVLQNVQOQKDPALLSLPTDRPRAPQTFAGAYHEFALSVELTKLT
KUSQGVYTFWMLLAADFLLRYTEQSDILVGTPIANNRSEIEGLIGFVNTLV
RTDLSGNSFABELLTVRLMAMDAYAHQDLPEMLVEVLEQPERDLSHAPLFOVDFLLQ
NDPLSTVELTGLTVSPLETESATKFDLTLMQNTGNGLGVWEYNTDLDLFDHSTIERM

gene

CDS

KGHVFMLEAIVANPOERIDOLPMLTAVEQOOLFVENDTQADYPODKCIHQLEFEQV
KFPSPDAVAVIFENQOLTYGELNTQANLAHLQSLGVRPELVYKILPERSISITVLL
AILKAGGYPVLPDIDYTOOKLADISEDSQSVLITQOKLLNSLIPVQGVKIVLDTSE
MLTNOSOVNPSVEKPNELACLILYTSGSTKPKGMVLTAAVLSNSAISSEVFLTS
DRVLOFASFSDFAAEELFTWYKATVLRPTQMFDFDLASQFIAQESLVNLTTP
AYWHEWAVSQDNATVPPSRLVAVGGDAVLLETITVIRQVGVDRVNCNLNVGNTIP
SVTAIVHDLHLPQSETANSVLIGRPIANTQAYILDRHLQPVPGVKGELHLCGVRLAR
GYNRPBELTEKEIPNPSGSGAEILANSQSPVSRSLYKTDGLARYLPDGNIECFG
RIDNQVKIRGFIEIGEIAVLMQYIDQVTSKVIDRDLPAEDLSSNRDQVYVPPET
ISELROFLSNPLVYMPQAFVLDPLPLTRKRVDRRALPADLSSNRDQVYVPPET
PTHEMLAHLVAEVEGAVGICHDFELGVHSHVQWRYWYVLPVLPVELPLSLFAAA
TGAQLADLIGOSQRNLEITVPLIPRAENAEILSTAQORLWFLDQLOPNSLYNIP
LAURLAFLVEAALQESLEIEIYHEALRTNFTIDGKPTQIITVNRNQLSVLEIHLQ
LPREQETSQELAITAQIOPFDLQSLRLATLVLSSETHLLVCHHIIVSDGWS
GVTOELATSYNAVQOSANTAPLIQYADFAINQRELQGLDLSQLSYQQQKOLKQ
APALLTPDRPRPAPVKSGLQAEFALSLELTQGLMOLSQOQGVTFEMTLLAGFDL
LYRYTGSDILVGTPIANNRSEIEGLIGFVNTLVNRSNLSNPFAVLSLIGRVENA
LSAYAHQDLPEMLVLEALQORNLSTHTPLFOVAVFQNPANPSQVLELTVSPLVES
Query Match 9.8%; Score 97.6; DB 1; Length 39436;
Best Local Similarity 51.5%; Pred. No. 7.8e-17;
Matches 279; Conservative 0; Mismatches 254; Indels 9; Gaps 2;
QY 58 CAAAGACCCGCTGTGACGACTCACTCTCTACCTATCCAGTACAGCAGCTTTGCAAAAT 117
DB 13392 CAGACGCTACGGAATACCTTTAACTCCACTGCGGATTCTAGCTACGAGATTTTGC 13451
QY 118 GGCAGAGGACCAATTCAT-----AGACGAGGAGAGCAACCACTCACTGGAAGAAC 171
DB 13452 GGCAGAGAACTGGTAGTCGGAGATGTACTGCAAGGCAACTGAATTAATGTCGCAAC 13511
QY 172 AACTCAAGA---CTCTTCCCCAGCAAGATCCGACCGACTTTGCCGCCCTGCACTTC 228
DB 13512 AACTAAAGATGCCCCAGCCCTGTAGTGTCTGCCACAGACCGACCGACCTGCTGTC 13571
QY 229 TGTCTGAGACGACGAGTGTGCTACATGTACCATCGACGCGAGCTCTACAGCTCCCTTC 288
DB 13572 AACTCTTGTAGCGGCATATCAAGATTTTCTTGTGGAATTAACCAAGGGTTAA 13631
QY 289 GAGCCTTCTGCAACGACCAACACGACCTCTTTCTGCTGTTCTTCTAGCTGCGCTGCG 348
DB 13632 TGCAACTGAGTCAGCAACAGATTTGTTACTCTGTTTATGACGCTGTTGGCAGCGTT 13691
QY 349 CGCTCATTTATCTCTACAGCTGTTGAGACGCTGTCTATGTTACACCACTTACGCAAT 408
DB 13692 CCTTCTTTTCTCTACAGGACAGACATTTTGTGAGGACTTCCAAATGCTAAC 13751
QY 409 GCAACGACCTGAACCTGGAGATATCATCGCTGCTTTGTGCAATACGCACTGTATCGAA 468
DB 13752 GCGATCGAGTGAATAGAGGGTTAATGGCTTTTGTGCAATACCTTTAGTGATCGGA 13811
QY 469 TCAACATAGATCATCAGTACCTTTGGGACTTTGTATCAACCACTCAAGGCTACGACGA 528
DB 13812 GTAATTTATCGGAAATCCAGTTTTCGCGAATTAATCTACTCTGTTTGGCTCTATG 13871
QY 529 CAGCACCATTCGAGACGAGGATATTCGCTTTGAGCGCGTTTGTATCAGCACTACAGCCT 588
DB 13872 TATCCGCTTATGCTCATCAGATTTTGCATTTGGAATGCTGTGGAAGCATTCAGC 13931
QY 589 GA 590
DB 13932 AA 13933

RESULT 11

AF455810

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

AF455810 15582 bp DNA linear BCT 02-JAN-2002
Xenorhabdus bovienii T228 ABC transport protein xpsd (xpsd) gene,
partial cds; and peptide synthetase operon, partial sequence.
AF455810
AF455810.1 GI:18034620
Xenorhabdus bovienii.

ORGANISM Xenorhabdus bovienii
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Xenorhabdus.
REFERENCE 1 (bases 1 to 15582)
AUTHORS Pinyon, R.A. and Thomas, C.J.
TITLE Identification of a novel peptide synthetase operon in Xenorhabdus bovienii T228
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 15582)
AUTHORS Pinyon, R.A. and Thomas, C.J.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Microbiology and Immunology, Adelaide University, Molecular Biosciences, Victoria Drive, Adelaide, SA 5005, Australia
FEATURES
source 1..15582
/organism="Xenorhabdus bovienii"
/strain="T228"
/db_xref="taxon:40576"
<1..724
/gene="xpsD"
<1..654
/gene="xpsD"
/codon_start=1
/transl_table=11
/product="ABC transport protein XpsD"
/protein_id="AAL57598.1"
/db_xref="GI:18034621"
/translation="IVHVTDKERQFMGLPLSLKISQGEIVFVGGNGSGKTKLAM
MLVGLPEQSGSLWLVGVNDASNNVHYRQFFSAVSNHLEFDLLNTGDDVTEKATH
YIOALNHPKRLIIDGKFTSTELAGORRKLALVAAYLEDRPLYLDFEWAADDPVFK
RLFYTNLLPELRSRGRTVIVISHDNAYFDIAERVIKLEDGNIKEINNHSHETITELK"
97..117
/gene="xpsD"
/note="Region: Walker A motif"
373..399
/gene="xpsD"
/note="Region: linker peptide"
433..456
/gene="xpsD"
/note="Region: Walker B motif"
547..549
/gene="xpsD"
/note="Region: invariant histidine residue"
701..724
/gene="xpsD"
988..>15582
/gene="peptide synthetase operon"
988..993
/gene="peptide synthetase operon"
1011..1016
/gene="peptide synthetase operon"
/note="putative"
1031..1036
/gene="peptide synthetase operon"
/note="putative"
1053..1058
/gene="peptide synthetase operon"
/note="putative"
1127..1140
/gene="xpsA"
1127..1133
/gene="xpsA"
1141..1140
/gene="xpsA"
/codon_start=1
/transl_table=11
/product="peptide synthetase xpsA"
/protein_id="AAL57599.1"
/db_xref="GI:18034622"
/translation="MNHPEKLKPFALSEAQSRRWFQYQINPSQGRNNGAFCAVRKGL
NAQDLEALNHLIKRPHMLRASFGLYDGLGYRIENSEIKVQTHDRHLSEDTLQQL

VFDDCHVLDLENLPRFASWYOCNAQESVLVTFDHLVVDGWSYWLLEELGMLTSS
CELEPTSEHSYQDYVSWQOQLNSKSAKQOQFCWQCNLGNLSLSPKIRYIGSVQNGV
DNNASPTITDKITAFGRIPDDLALKLCMSAEXGNSLFAIFLSAYQILLNRYAQDDI
VIGSMMPGRSQAQKGLVGEFVNPVLRGQINGNTVQEHILQITAKTIRQAIENQRYP
F7KVLQFQKLORTADTHPVQTLMTQKPRYVSELPTLWDDDSSTVHGGAELEP
PHLYADTPVPLMNMMEVNNQQRCHFHYPQVDFDAELINQLIKNLITLITANADNQ
QSIDSPLMDEQEQHQQVLYDFNKTPFPQHALIHQLEQQAERTPAIALSCGDKAL
SYTELNQANRLAHALIKAGIRPDNRVAICMERSLEMVIGLLGILKAGAAAYVPLDPEY
PDLRLGNILSDSPALILIHHGLQDHLPMTPVWVLESEYRTNIAESQPTDNPVAD
LGLTSRLHVLVLTSGTGLPKGMNEHGVNRLWAOEYOLTOHDRVLOKTPPSE
DVSWEFFELPLLAGTOLVMPKPGHKEALYLLEIEARGITTLHFVPSMLQSTHLTP
AGRPSLRQLCSGEALSYSQQCLAHFAHSELHNLGYFTEAAILNDRVCPDQHI
GLVPIGHPIIDNTQIYILDKHQDPVIGVIEIYAGAVARYLNDKPELTAERFIRD
FSQHPDMRYKTGDIGRWLADSGIDYLRNDFQVIRGNRIELGETEARLAQSDGVQ
VIVITAREYDAGDRLVAYLIPKPGVTLSIPALREQVGGSLPYDMPISAFVMLDAFPLT
LNGLDKRKALPAPDHSAVLTREYAAPQGETEQLADIWQKLIKIDRVGRNDFELGG
HSLLMLQLQSRISENFDVELSIQQLFAHPTICQLEEHIIIDQLLOPDAESLQDLYKTM
G"
1174..4239
/gene="xpsA"
/note="Region: XpsA module 1"
1177..1206
/gene="xpsA"
/note="Region: module 1 condensation domain C1 motif"
1315..1341
/gene="xpsA"
/note="Region: module 1 condensation domain C2 motif"
1549..1578
/gene="xpsA"
/note="Region: module 1 condensation domain C3 motif"
1651..1671
/gene="xpsA"
/note="Region: module 1 condensation domain C4 motif"
2020..2052
/gene="xpsA"
/note="Region: module 1 condensation domain C5 motif"
2128..2151
/gene="xpsA"
/note="Region: module 1 condensation domain C6 motif"
2293..2316
/gene="xpsA"
/note="Region: module 1 condensation domain C7 motif"
2662..2679
/gene="xpsA"
/note="Region: module 1 adenylation domain A1 motif"
2797..2829
/gene="xpsA"
/note="Region: module 1 adenylation domain A2 motif"
3034..3081
/gene="xpsA"
/note="Region: module 1 adenylation domain A3 motif"
3184..3195
/gene="xpsA"
/note="Region: module 1 adenylation domain A4 motif"
3466..3486
/gene="xpsA"
/note="Region: module 1 adenylation domain A5 motif"
3622..3666
/gene="xpsA"
/note="Region: module 1 adenylation domain A6 motif"
3736..3753
/gene="xpsA"
/note="Region: module 1 adenylation domain A7 motif"
3790..3849
/gene="xpsA"
/note="Region: module 1 adenylation domain A8 motif"
4000..4020
/gene="xpsA"
/note="Region: module 1 adenylation domain A9 motif"
4057..4077
/gene="xpsA"
/note="Region: module 1 adenylation domain A10 motif"
4207..4239

```
/gene="xpsa"
/note="Region: module 1 thiolation domain motif"
4448..14398
/gene="xpsb"
4448..14398
/gene="xpsc"
/codon_start=1
/transl_table=11
/product="peptide synthetase Xpsb"
/protein_id="AAL57600.1"
/db_xref="GI:18034623"
/transl_table="MNDNELISPLAERKRLLEAKAKLTKROCTOKTEIHAOPDCN
VPLSWAQRLEFQLDPAQAQAYHMSAGNLQHLNQNALKAALDOIVARHEILRTT
IVDEGQOQIGDSAGFSLVSDPLPSTEQQAEECAEQADAEPLPALTLQYLA
RGRLLRAEHSVLLLTQHHIISDGWSNIMLSELTLYAQFQDQAEPLPALTLQYLA
DYALQWQGLQVLEQDQWRSELAGAPVLELDPKPRTOQSQVAGSRHELEPLIF
ALSTELKAFSQRHGTITFTLLAGWALLSRISGQHLVIGSPVANRQSRHELEPLIF
FVNTLALRIQDNPSSVELLARVKHALGAYAHODLPEQLVEALPPBSLGHSPIF
QVMLALDNTGQOYFELDGLHLLHELPTROSAYFDLTLTNDTEQSLVGLDEYASDLF
EHASIERWVGLTILSAMVADSLRVDLPLTSSORTQLLANFNDAIPIPKNALI
HQLFEQVETPKIALVWGETQLSYSELNQRANQLAHSMASGVHPDDRVAICAERS
LDMVIGVGLTKAGASVPIPLDNPHTERLAYMLSDQFVLMLTQOHLKARLPVTNIVP
WALDSEHQCTPACQDNIDASQGLTSLNAYLVLTSGTGLPKGMTEHQNVHL
IHSQFOMSKLTPHCVLQFASEFGDNSVAEFTPTFAIGATVVLPRDHIKYPDTEFITF
LONGIVLPLPTAFWHLMAQELISAGYSWPEQLRSLVAAGEKAHRHLTWLSSPT
QKRWLNTGPTTETVWATSIIVDKENLYEDIPGRPIANTRIYILDRGQDPVPIG
VNGEIHSSGVARGYLNRSELTAEFIQDPFDPGARMYKTGDLGRPDFTISYS
GRNDFQVIRGFRELGEIEAQLATCAQKDAVIVREDNDGKRLVAYLTIQSGAIL
NAASRLQSLVNDAYMLPSAVFTLEAFPLNQNGIKIDRPALPAPDRASVREYAEQ
GEIQQOALQWNLGRICRYDSFELGHSLLTVQVARSRLQSLNIETIQLDFT
QPVLDLAQSLVATOSTOPMILPAREQALPLSWTOORLWFLAQLDPAQATVNMGS
GLHQLQHLNQNALKAALDRIVARHEILRTTIVQAEKARQVIGDACGFSLTVRDLQ
LSSQAQCTIEECAFASHPDFAQGLIRAQKLLKEQOQHVLLLTQHHISDGWSL
NVMNDELISPLAERKRLLEAKAKLTKROCTOKTEIHAOPDCN
APTLLEPTDKVRESVQSGHGDQVFTLSPENSLGRALSORHGATFTVLLAGWGL
LSRLSGQPLDIVGPVANROYSELEPLIGFANTLALRIKLENDPTVSALLARVKAH
LEAYAHODLPEQLVEALPPBSLGHSPIFQVMLALDNTSSKQSFELAEISLNPALT
RNSAHPDLTLALSDENSLCELYASDLPEPERSIERMAGYLNQLLAAMVADNLRVA
DLPLMHPTERTLTDNDTAVYPODKLSQLFQEQVETPKIALVWGETQLSYSELN
LQNRALQHALTAFFGQDPRDRAICIERNLNVMVIGMLKAGAGVYDPDPEPAER
LAYTLSDASAPLTLQOHLQALPEKVLQWQDGLHNSVAQDNDPDRGLGF
PHLAYITVTSGLTGLPKGMTEHQNVNFTYSQCTSELKSTDRVLFQFASPDFTAV
SEIPLTSLTLRLPAHRIIPDATESHFLQEQALISVIDLPTAFWQHVQEMKAGVS
GFSSHVSIRVTVGGKAEHRHFTWQSMPTRHCHWIDTYGPTTETVTSATLTLDGPAS
YVYESLSIGRPLINTHYVILDTQGPVPIGVTEGIYIGGAGVARGYLNRBELTAERFV
SDPREQPHARMYKTGDLGRPDFTIIVLGRNDFQVIRGFRELGEIEQSLAAGTG
VSDAVVAVAREEGSGDKRLVAYLPKSDVTLDAASLRQVSTHAEYMLPSAFVILDAF
```

Query Match
Best Local Similarity 49.18; Pred. No. 5.1e-14;
Matches 292; Conservative 0; Mismatches 297; Indels 6; Gaps 2;

```
Qy 4 TTGATGCTTCGCGCGGATCTCAATCAGCTCTACTCAGCTGCGCTCAAGGACTCAAAAG 63
Db 8232 TTAATGATTGATGAACGAATTTCTCGCTGTACCAAGCTTCGGTCAGGGCAAGCTG 8291
Qy 64 ACCGCTGTGAGCACTCACTCTACCTATCCAGTACAGGACTTTCGAAAATGGCAGA 123
Db 8292 ATCCATTGCGCTGATGACCAATTCATATGCGGACTA---TGCGGTTTGGCAACGGCAAT 8348
Qy 124 AGGACCAATTCATAGCAGGAGGAAGCAACTCAACTACTG---AAGAACCAACTCAAG 180
Db 8349 GCGTGAAGGGGAAGGCTGAAAACAGTTGAGTTACTGCGGCAATGAGTTACGAATG 8408
Qy 181 ACTCTTCCCAGCAAGATCCGACCGACTTTGCCCGCCCTGCACTTCTGTCTGGAGACG 240
Db 8409 CCCCACCTTATTAGACTGCTACTGATAAAGTCCCGCTCAGTACAGAGTTATCATG 8468
Qy 241 CAGGTCGCTACATGTTTACCATCAGCGGAGCTCTACAGTCCCTTCGAGCCCTTCGCA 300
Db 8469 GTGATCAGGTCACGCTTCACTTATTCGCGCGGAATTAATTCAGGGTTAAGGGCAATTGACC 8528
Qy 301 ACGAACACACACGACGACCTCTTCTCGTCTTCTTCTAGCTGCTTCCGTCGCGCTCATTC 360
```

8.88; Score 87.8; DB 1; Length 15582;
49.18; Pred. No. 5.1e-14;
Matches 292; Conservative 0; Mismatches 297; Indels 6; Gaps 2;

Qy 4 TTGATGCTTCGCGCGGATCTCAATCAGCTCTACTCAGCTGCGCTCAAGGACTCAAAAG 63
Db 8232 TTAATGATTGATGAACGAATTTCTCGCTGTACCAAGCTTCGGTCAGGGCAAGCTG 8291
Qy 64 ACCGCTGTGAGCACTCACTCTACCTATCCAGTACAGGACTTTCGAAAATGGCAGA 123
Db 8292 ATCCATTGCGCTGATGACCAATTCATATGCGGACTA---TGCGGTTTGGCAACGGCAAT 8348
Qy 124 AGGACCAATTCATAGCAGGAGGAAGCAACTCAACTACTG---AAGAACCAACTCAAG 180
Db 8349 GCGTGAAGGGGAAGGCTGAAAACAGTTGAGTTACTGCGGCAATGAGTTACGAATG 8408
Qy 181 ACTCTTCCCAGCAAGATCCGACCGACTTTGCCCGCCCTGCACTTCTGTCTGGAGACG 240
Db 8409 CCCCACCTTATTAGACTGCTACTGATAAAGTCCCGCTCAGTACAGAGTTATCATG 8468
Qy 241 CAGGTCGCTACATGTTTACCATCAGCGGAGCTCTACAGTCCCTTCGAGCCCTTCGCA 300
Db 8469 GTGATCAGGTCACGCTTCACTTATTCGCGCGGAATTAATTCAGGGTTAAGGGCAATTGACC 8528
Qy 301 ACGAACACACACGACGACCTCTTCTCGTCTTCTTCTAGCTGCTTCCGTCGCGCTCATTC 360

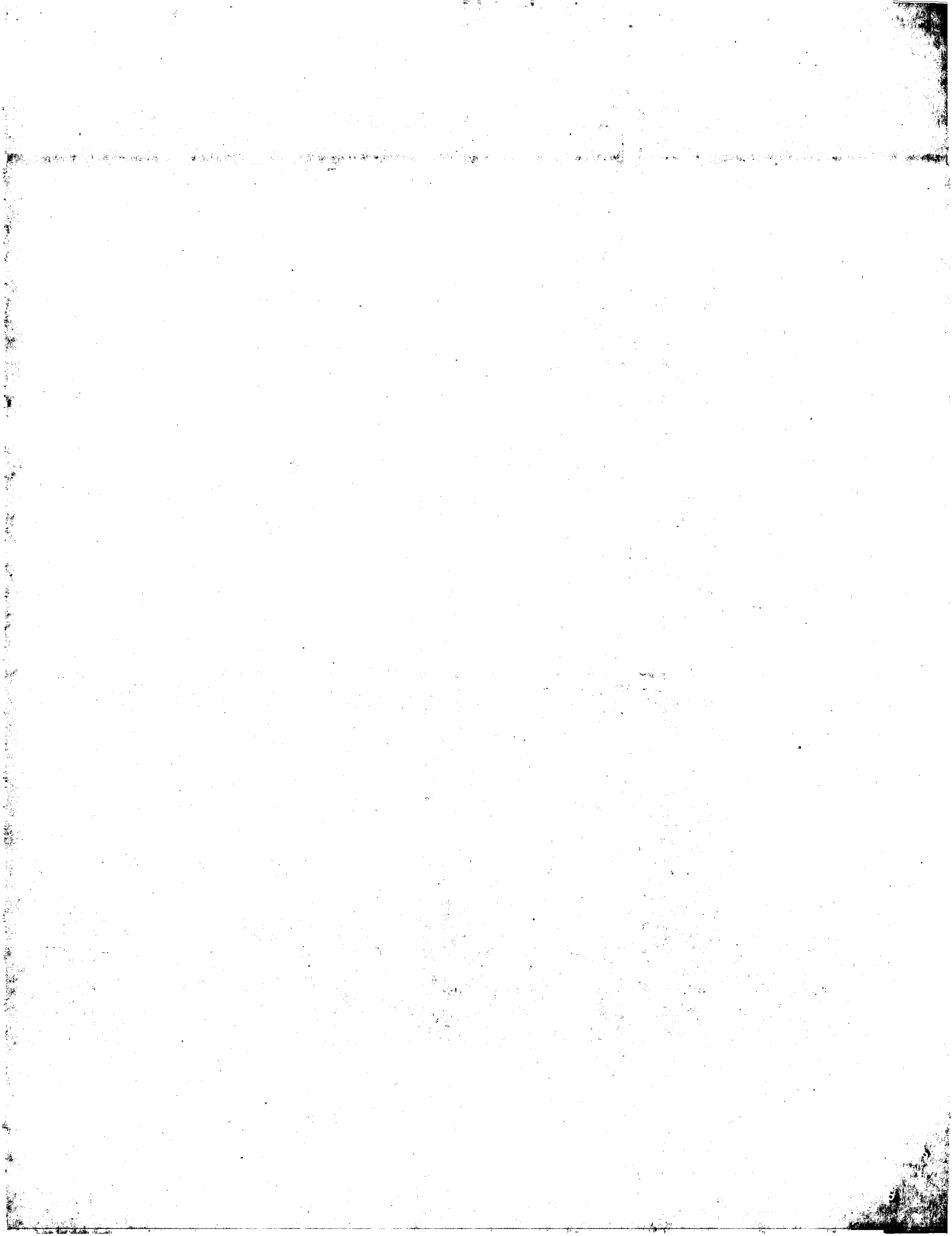
```
Db 8529 AGCGCCACGGTGGCGACGTTGTTTCATGACATTACTGGCGGCTGGGGAATTTTGTGTCTAC 8588
Qy 361 GTCTCACAGCTGTTGAAGACGCTGTCTATTGTTGACACCAATTTGCGAATCGCAACCGACCTG 420
Db 8589 GGTAAAGTGGTCAGCCCGATTTAGTATTGTTACTCCCGTCGCCAACCAGCCGATACAGCG 8648
Qy 421 AACTGGAGGATPATCATCGGCTGCTTTGTGTCATACGACGATGTATGCGAATCAACATAGATC 480
Db 8649 AGTTGGAGCCCTTAAATCGCTTCTTGGCCATATACACTGGCCCTGAGAAATCAAGTTAGAAG 8708
Qy 481 ATCAGGATACCTTTGGGACTTTGATCAACCAAGTCAAGGCTACGACGACGACGACGATTCG 540
Db 8709 ATAACCCGACGCTGAGCCGATTTATGGCAAGAGTGAAGGCTCATGCGCTGGAAGGCTACG 8768
Qy 541 AGAACGAGGATATTCCGTTTTCGAGCGCTTTGTTATCAGCACTACAGCTCGGATCCAG 595
Db 8769 CCCATCAGGATCTGCCATTTGAACAATTAGTGAAGTCTACAAACCGCCACGCGAG 8823
```

RESULT 12
PSPBSC
LOCUS PSPBSC 3671 bp DNA linear BCT 03-AUG-1994
DEFINITION Pseudomonas sp. pbsC gene.
ACCESSION X77699
VERSION X77699.1 GI:455513
KEYWORDS pbsC gene; protein C.
SOURCE Pseudomonas sp.
ORGANISM Bacteria; Proteobacteria.
REFERENCE 1 (bases 1 to 3671)
AUTHORS Adams,C., Dowling,D.N., O'Sullivan,D.J. and O'Gara,F.
TITLE Isolation of a gene (pbsC) required for siderophore biosynthesis in fluorescent Pseudomonas sp. strain M114
JOURNAL Mol. Gen. Genet. 243 (5), 515-524 (1994)
MEDLINE 94268496
PubMed 8208243
REFERENCE 2 (bases 1 to 3671)
AUTHORS Adams,C.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1994) C. Adams, University College Cork, Microbiology Department, Cork, IRELAND
FEATURES
source 1..3671
/organism="Pseudomonas sp."
/strain="M114"
/db_xref="taxon:306"
1089..3546
/gene="pbsC"
1089..1099
/gene="pbsC"
1135..3546
/gene="pbsC"
/codon_start=1
/transl_table=11
/product="biosynthetic protein C"
/protein_id="CAA54778.1"
/db_xref="GI:455514"
/db_xref="SPTREMBL:Q52531"

/transl_table="MIEELVLYAGFNLDHAGLPALPIOVADYALMQRORWEAGEQE
ROLAYWQEKLGGEOPVLELPMDRSRTIEQSFAGASHNLLILDPAISNALKAFARQENVT
LFVLLASQALLHRYSGQADIRVGVPNANRVEIERLIGFFVNTQVLRADIDGOMS
FVLLDRVKTAQQAQAHQDLPEQLVEALEPGLSHSPLFQVFNHQAERRASVET
RLNGLSIKPLEQSQTAQFDLTLNTEQAHEIAVLKYATDLDLDAATIERLAQHNAL
LQALVAAPTQRIQGLPMDAQOQLLAQWNPVERVEQPLEQICHOATEAQAERYPDAI
AVTYDGRLTVAELNCRANQWAVLIARGVPDVRVGVAVERSMDMIVAIIAIVKAGG
AYVPLDPCYDDRLSYMIADSGIELLTQGHLLAQLVPVPSGLACLDIHOPEPNETN
PLCLTTPDLSLAYVIYTSCTGSKPKGALLPHGNVRLFSATEHWFDFQDQSWTLFHSY
AFDFSWVEIFGALLYVGLKLVVPHDLSRSPEDFYLLICDENVTNLTNPSPAFKLMQV
ASESTRNNSRLGVVFGGLEVLSREWFEDGAPLTINMYGITTTVHTVYRPLS
MADEQSHSPNIGPEIVDLNLSVLDGSLNVPQCCIGELYTAGARGYLNQAGMTA
TRVPDPPNQAERLYRTGDLARLDGVIIEYIGRIDHOVKIRGRFELCEIEAQLL
KHADVREAVVLAVDGISGHATGDVDCRERRTGRRAAGTARFDDQAGARNPAGLHGS

| | | | | | |
|---------------------------|---|--|--------|-------|------|
| BASE COUNT | 717 a | 1128 c | 1135 g | 691 t | gene |
| ORIGIN | QLGVP" | | | | CDS |
| Query Match | 8.6%; Score 86.2; DB 1; Length 3671; | | | | |
| Best Local Similarity | 49.7%; Pred. No. 1.3e-13; | | | | |
| Matches 248; Conservative | 0; Mismatches 248; Indels 3; Gaps 1; | | | | |
| QY | 100 | ACAGCGACTTTGCAAAATGGCAGAGGACCAATTCATAGAGCAGAGAGCAACTCAACT | 159 | | |
| Db | 1220 | ACTAGCGCTTATGGCAGCGGAGTGATGGAGCGGTCGAACAGAGCGGCAACTGGGCT | 1279 | | |
| QY | 160 | ACTGGAAGAGCAACTCAAGACTCTTCCCCAG---CAAAGATCCGACCGACTTTGCC | 216 | | |
| Db | 1280 | ACTGCGAGAGAAACTCGTGCGGCAACAAACCGGTGCTGGAATCGCGATCGCTCGC | 1339 | | |
| QY | 217 | GCCCTGCACTTCTGCTGAGACGCGAGTTCGCTACATGTTACCATCGACGCGGAGCTCT | 276 | | |
| Db | 1340 | GCACCATCGAGCAAAAGCTTCGCTGGCGCCAGCCATAACTGATTCTCGACCGCGCTGA | 1399 | | |
| QY | 277 | ACCACTCGCTTGCAGCCTTCTGCAACGACGACACACAGCACTCTTTCGTCCTTCAG | 336 | | |
| Db | 1400 | GCAACGCGCTCAAGACCTTCGCGCGCAAGAAAGCTCACCCCTGTTCTGCTGCTCGT | 1459 | | |
| QY | 337 | CTGCGCTTCGCTCGCTCATTATCTCTCACAGCTGTTGAACAGCGCTGTCATTGCTACAC | 396 | | |
| Db | 1460 | CCTCGTTCAGCGCTTGTTCATCGCTACAGCGGTCAGCGCCAGCATCCGCTCGCGCTGC | 1519 | | |
| QY | 397 | CAATTGCGAATCGCAACCGACCTGAACTGGAGGATATCATCGGCTGCTTTGTCAATAGC | 456 | | |
| Db | 1520 | CGAATGCCAACCGTGGTGGGTAGAGATCGACGCGCTGATCGCTTCCTCGTCAACACTC | 1579 | | |
| QY | 457 | AGTGTATGCGAATCAACATAGATCATCAGGATACCTTTGGGACTTTGATCAACCAAGTCA | 516 | | |
| Db | 1580 | AGGTGCTCAAGCGCGACATCGATGGCGAGATCTCTTTGTCGATGCTGCTCGCTCAGGTC | 1639 | | |
| QY | 517 | AGGCTACGACGACAGCAGCTCGAGACGAGGATATTCGCTTTGAGCGGCTGTGATCAG | 576 | | |
| Db | 1640 | ACACACCGCGCAAGACGCGAGCGCCATCAGACTTCGCGTTCGAGCAACTGGTCGAAG | 1699 | | |
| QY | 577 | CACACGCTGGATCCAG 595 | | | |
| Db | 1700 | CGCTGGAACCGCGCGCAG 1718 | | | |
| RESULT 13 | AF047828 28587 bp DNA linear BCT 01-DEC-1998 | | | | |
| LOCUS | Pseudomonas syringae pv. syringae syringomycin synthetase (syre) | | | | |
| DEFINITION | gene, complete cds. | | | | |
| ACCESSION | AF047828 | | | | |
| VERSION | AF047828.1 GI:3510628 | | | | |
| KEYWORDS | Pseudomonas syringae pv. syringae. | | | | |
| SOURCE | Pseudomonas syringae pv. syringae | | | | |
| ORGANISM | Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; | | | | |
| REFERENCE | 1 (bases 1 to 28587) | | | | |
| AUTHORS | Guenzi,E., Galli,G., Grgurina,I., Gross,D.C. and Grandi,G. | | | | |
| TITLE | Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases | | | | |
| JOURNAL | J. Biol. Chem. 273 (49), 32857-32863 (1998) | | | | |
| MEDLINE | 99047670 | | | | |
| PUBMED | 9830033 | | | | |
| REFERENCE | 2 (bases 1 to 28587) | | | | |
| AUTHORS | Guenzi,E. and Grandi,G. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (12-FEB-1998) Mol. Biology, Chiron/Vaccines, Via Fiorentina, 1, Siena 53100, Italy | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..28587 | | | | |
| | /organism="Pseudomonas syringae pv. syringae" | | | | |
| | /db_xref="taxon:321" | | | | |

/note="pathovar: syringae"
285..28415
/gene="syre"
285..28415
/gene="syre"
/note="Syre; peptide synthetase"
/codon_start=1
/transl_table=11
/product="syringomycin synthetase"
/protein_id="AAC80285.1"
/db_xref="GI:3510629"
translation="MGQRLSKAPFKLEVTDFPEVDGDTQEAIAHLRESFPQFESV
TGLMESLVRGCPQLHFWNRLHVLADGICAVLIAHAYSDAYSGLLEKGTLEAGP
SYLSLEEDRAYMASSRYKRDRLFWEQSYAQLPPVLLQRRADFAGLNTNPPASQVQ
AMLPRALYNALQFASACNLSVAHLVGYISTYFCFTGVDEIVVGMVPHNTRAQK
ATIGMFSSPIRVGDFPHETLDLMNIGTQLRRTYRHQREPIAEINLSRLAONGR
HOLFVLSFESFDGVTGSDPDLMDNGYEOTPMAFIVRDYHPSEDTHLDFNFR
NTAYFTFEAQRVQORIPAMLEAVLNQDTPLAHFLPMSTSEHQALATFNDTAHYPR
DYLIIHOLIEQQAQRDACAVRGDSGPLTYAELNQANQLAHLRIELGVEYEDTRVAV
SURGAEMVALLGILKAGGATVPIDPLPSARQAIMLEDSFQAVLTITROLDNLPA
SDPLVLGDHNDRAQLARQQSNPDAKALQPNHLAYLVYSTGSTGTPKGMNEHL
GYVNRLLWARDAYQVNSQDRVLQKTCGFDLSWEEFFLLPTGAELMAPPGHQDPP
YLAQVMSDAGITLLHFVPSMLDVFLEHRSRDFPQLRRVLCSEALPRALQRRFEOQL
KGVLEHNLGPTTEAALDYTAWECPRTDGDSPVIGRPIANIOMHVLDAQLQPMGVA
CELHTGGVARGYLNQOLSARETIADPFSNDQARLYKTGDVGRWLANGALEVLR
NDFQVKIIGRLRIEIGIEIAALAKHFAVHEAVTAREDIPGKRLVAYITQSEHTAVD
LEALSHLQQLPEYMPVPAIYVLLAMPITNSGKDKRKALPAPDGDALSYEAPQG
EYETLLASIWADLVKEQVRHDFEELGSHSLAVKJTERMQVGLSADYHVLFGQP
TVAAALAAAGGSEVVVPANLIPHECHDRITPQMLPLINLSQAQIDHVKMDPGGVANV
QDIYPLAPLAGILYHHSAEQGDPTLKALFALSDAQLODFSGALQGVINRHDILR
TAVLMEDDEPVQVVLROAELHVTLLDPVDGPDVEQLHQBDRHRHVLDTAPLM
RIVFSDPVDNRWLAMLCHLHLVSDATLSVLTQELIQAHLQCGNALGAVPIRNVVA
QARQGVSAQHEAFFRDMIGDIDEPTLPFGLOVDQDGRNLEEA SVTLAAELNLRRA
QALQGVSSVAHMLAWRVLGNVSACEQVFGTVLLGRMQAGDADRALGMFINLTP
LRVDIGATTVAEGLKATHERLTALLGHEHAPLVLAQRCGVAAPTPLFSALLNHRYSV
VSDVRKLPGLGSSLRHRRPYQLVDLNVDDLGAFTLTIOAQOQISATRIGETYMOV
ALRNVLDALEHTPOALNSLSILPDERELLITGFNDTAHYPRDVLIIHOLIEQQAQ
RPDACAVRGDSGLTYAELNQANQLAHLRIELGVEYEDTRVAVLSRGAEMVALLG
ALLKQPTGNDPAKALDQPNHLAYLVYSTGSTGTPKGMNEHLGVNRLNWARDAYQ
VNSQDRVLQKTPGFDVSWEEFFLLPTGAELVMAVPSGHQDPTLAQVLSIDAGITLL
HFVPSMLDVFLEHRSRDFPQLRRVLCSEALPRALQRRFEOHLEHNLGVPTEA
ADVTAWECRPTDGDSPVIGRPIANIOMHVLDAQLQPMGVACELHTGGVARGY
LNQOLSARETIADPFSNDQARLYKTGDVGRWLANGALEYLRNDQFQVKIIGRLRIE
IIEAALAKHFAVHEAVTAREDIPGKRLVAYITQSEHTAVDLETKHQHQQVLP
YMPVPAIYVLLAMPITNSGKDKRKALPAPDGDALISRGYEAQGETEIQIWDLL
GVEQVRHDFEELGSHSLAVSLRQKQDLNNTVFTAPSVREMAAISQEKQ
ALFRAPANRIPAHCTQLTDPMLPLVLSVAQLIEATASAVPGGAANTQIYPLAQEG
ILFHYLLNERDAYLVRSTIEFDSRARDLAFLEGQTVIDRDHDLVSRSSVHWGLQVAV
QVHVQAQSPHITLIRPEDEALSOLDRLSDPGRLLRLDRQAPLLAYLARPDSERW
LUALIDHHMSDHTLELLEELILLMRGQSAELPPQPYREFVAQTLPASSPSAHEY
FTGRLADVDSPTAPFELLEVOGQNDVEESALSDLCARITQARERGMSPAVLFH
VAAQVLAARCTGRDDVFTGTAVTGRQTLGAERAMGMFMTLPVRVQALTSQVLEL
MATHRDLSELLSHEQASLALAQCSVATGVPLFSLNLRHQNDSQLQWGLRLD
SAERTNYPCLSVNDYQSDGLLHISVQADPQRCLAMMOCALEDTALAHTPKQEV
TOLDVLPAAERNLLETFNQTQDPTDLCIQHLEFAOVRTOPDAIYAVAFQARLSYA
ELNRQANRLAHLHILGIGCPDRAVATCIGVEMVMVGLLVKAGAAVPLDPAPAE
RLAYMIEDSQPALMTQRLHETLPTLTVLLDDDDQKRTETERDNDNPPVVALGVN
LAYVYIGSTGTPKGMVIEHRLGVNSVDAARLFDLSPTDTVQTNLQNFDSVEEI
FPALLAGATLTPSREIFGSEGTENHGINPTVHLTAHHHTLVAEHWKQPOVAEORLQ
HVLNVTGDALSQKLLKWDREPAHTRILINTYGTETATVSTCYTAHVAADAAESG
SCNATIGKPMANTRYLLDHAHQPPYVAGETIYIGGQGVARGYINLNEVNAERFLAD
PFSPSPDARMYKTGLARYMADGRIEYLRNDQFQVKVGRFRIEIEIARLNCITGVK
EAVVJAREDNPGDKRLVAVVVAQSQSLTAADRLAEAPQLAEIIMPASFVLLDESLPL
TPNRKLDKRALPADADALISRGYEAQGETEIQIWDLLGVEQVRHDFEEL
GSHLSLALIERLKRQGLNKNVFTVPTREMAALISQDKHVLFPVPAHTRIPAHCT
QLTDPMLPLVLSVAQIELITSVAGGAANIOTIYPLAPLQDGLFHYLNLRERDAYL
MRSLIEFDSRARDLAFLEGQTVIDRDHILSRSSVHWGLQVQVHVQAQPLPVHTLT
LTPEDALSOLDRLSDPGRLLRLDRQAPLLAYIARDPNSERWLLALIDHHMSDHT
VELILLEIRLLMRGQSAELPPQPYREFVAQTLPASSPSAHEYFTGRLADVDSPTAPF
ELVEQGDQNDVEEAKLALNSDCIRITQARERGMSPAVLFHVAQVLAARCTGRDD
VVFQTVTGRQGTAGARMGCMFMTLPVRVOLTQGAQELVMAHQAQVMACTGRDD
ASLALAQRCSSVATGVPLFSSLLNRYHQEDNRNLQWPGMRLLDGTERTNYPLCLSVND



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:28:42 ; Search time 271.5 Seconds
(without alignments)
8302.948 Million cell updates/sec

Title: US-09-482-788-1_COPY_4000_5000
Perfect score: 1001
Sequence: 1 caattgatgtcttgagcgc.....cccaagtctgtgtacc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

| Database : | N_Geneseq_101002:* |
|------------|---|
| 1: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:* |
| 2: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:* |
| 3: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:* |
| 4: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:* |
| 5: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:* |
| 6: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:* |
| 7: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:* |
| 8: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:* |
| 9: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:* |
| 10: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:* |
| 11: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:* |
| 12: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:* |
| 13: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:* |
| 14: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:* |
| 15: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:* |
| 16: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:* |
| 17: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:* |
| 18: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:* |
| 19: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:* |
| 20: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:* |
| 21: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:* |
| 22: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:* |
| 23: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:* |
| 24: | /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | ID | Description |
|------------|-------------|-------|--------|----------|--------------------|
| 1 | 100.0 | 11212 | 21 | AAA58762 | DNA encoding a cyc |
| 2 | 39.2 | 9633 | 22 | AAF79702 | Mycella sterilia c |
| 3 | 339.4 | 33.9 | 46899 | AAQ54386 | T. niveum Cyclospo |
| c 4 | 168.6 | 16.8 | 3973 | AAQ78276 | T. niveum alanine- |
| c 5 | 168.6 | 16.8 | 3973 | AAQ78281 | T. niveum alanine- |
| 6 | 85.6 | 8.6 | 37856 | AAA11992 | S. cellulosum DNA |
| 7 | 75.6 | 7.6 | 13029 | AA551470 | Pseudomonas aerugi |
| 8 | 75.4 | 7.5 | 7335 | AAF81367 | Quorum sensing con |
| 9 | 74.2 | 7.4 | 18660 | AAA58472 | Nucleotide sequenc |

| | | | | | | |
|------|------|-----|---------|----|----------|--------------------|
| 10 | 71.8 | 7.2 | 3471 | 24 | ABK74876 | Bacillus lichenifo |
| 11 | 70.4 | 7.0 | 5451 | 22 | AAF26319 | Pseudomonas sp lip |
| 12 | 67.4 | 6.7 | 77536 | 21 | AAA14651 | Nucleotide sequenc |
| 13 | 65.8 | 6.6 | 88421 | 24 | AA40781 | 88421nt genomic DN |
| 14 | 64.2 | 6.4 | 41599 | 22 | AAI68165 | Bacillus subtilis |
| 15 | 62.8 | 6.3 | 178 | 15 | AAQ54388 | T. geodes Cyclospo |
| c 16 | 60.4 | 6.0 | 7347 | 23 | AA554136 | Pseudomonas aerugi |
| c 17 | 60.4 | 6.0 | 4403765 | 22 | AAI99683 | Mycobacterium tube |
| c 18 | 60.4 | 6.0 | 4411529 | 22 | AAI99682 | Mycobacterium tube |
| 19 | 59.8 | 6.0 | 23666 | 12 | AAQ10190 | Cephalosporin anti |
| 20 | 56 | 5.6 | 7158 | 24 | ABK74880 | Bacillus lichenifo |
| c 21 | 52.4 | 5.2 | 4620 | 22 | AAF26318 | Pseudomonas sp lip |
| 22 | 52.4 | 5.2 | 31122 | 14 | AAQ40706 | Bacillus subtilis |
| 23 | 51.8 | 5.2 | 8301 | 22 | AAF90035 | Nucleotide sequenc |
| 24 | 51.8 | 5.2 | 34071 | 22 | AAF90033 | Nucleotide sequenc |
| c 25 | 51.8 | 5.2 | 42717 | 22 | AAF90032 | Nucleotide sequenc |
| c 26 | 48.6 | 4.9 | 58857 | 21 | AA58471 | Quorum sensing con |
| 27 | 48.4 | 4.8 | 2556 | 22 | AAF81368 | Bacillus lichenifo |
| 28 | 47.2 | 4.7 | 6465 | 24 | ABK74875 | Bacillus lichenifo |
| 29 | 47.2 | 4.7 | 8268 | 24 | ABK74878 | Bacillus lichenifo |
| 30 | 47.2 | 4.7 | 4403765 | 22 | AAI99683 | Mycobacterium tube |
| 31 | 47.2 | 4.7 | 4411529 | 22 | AAI99682 | Mycobacterium tube |
| 32 | 46.2 | 4.6 | 1954 | 22 | ABA89199 | Escherichia coli p |
| c 33 | 46.2 | 4.6 | 48715 | 22 | ABA89188 | Escherichia coli p |
| 34 | 41.2 | 4.1 | 11601 | 12 | AAQ13608 | ACV synthetase gen |
| 35 | 41.2 | 4.1 | 13058 | 14 | AAQ48231 | Vector containing |
| 36 | 39.8 | 4.0 | 4615 | 22 | AAF90034 | Nucleotide sequenc |
| 37 | 38.6 | 3.9 | 7178 | 21 | AA59145 | DNA encoding a pep |
| c 38 | 36.8 | 3.7 | 1038602 | 20 | AAZ01425 | Complete genome se |
| c 39 | 36.2 | 3.6 | 273254 | 21 | AAC81914 | Chlamydia pneumoni |
| 40 | 36.2 | 3.6 | 1230025 | 20 | AA91990 | Nucleotide sequenc |
| 41 | 35 | 3.5 | 4590 | 22 | AAH24065 | Yeast AOB9604-asso |
| c 42 | 34.8 | 3.5 | 3519 | 23 | AA581987 | DNA encoding novel |
| c 43 | 34.8 | 3.5 | 3587 | 23 | ABL20034 | Drosophila melanog |
| c 44 | 34.8 | 3.5 | 3501 | 23 | ABL18792 | Drosophila melanog |
| 45 | 34.8 | 3.5 | 4024 | 20 | AA84595 | MTG16b protein cod |

ALIGNMENTS

| | | |
|----------|--|--|
| RESULT 1 | AAA58762 | DNA encoding a cyclohexadepsipeptide synthetase. |
| ID | AAA58762 standard; DNA; 11212 BP. | |
| AC | AAA58762; | |
| XX | | |
| XX | | |
| DT | 20-OCT-2000: (first entry) | |
| XX | | |
| DE | DNA encoding a cyclohexadepsipeptide synthetase. | |
| XX | | |
| KW | Cyclohexadepsipeptide synthetase; filamentous fungal cell; | |
| KW | Cyclohexadepsipeptide; antibiotic; ss. | |
| XX | | |
| OS | Fusarium venenatum. | |
| XX | | |
| FH | Key | Location/Qualifiers |
| CDS | | 270..9659 |
| FT | | /tag= a |
| FT | | /product= "cyclohexadepsipeptide synthetase" |
| XX | | |
| PN | WO200042203-A2. | |
| PD | 20-JUL-2000. | |
| XX | | |
| PF | 13-JAN-2000; 2000WO-US00913. | |
| XX | | |
| PR | 13-JAN-1999; 99US-0229862. | |
| XX | | |
| PA | (NOVO) NOVO NORDISK BIOTECH INC. | |
| XX | | |
| PI | Berka RM, Rey MW, Yoder WT; | |
| XX | | |

| | | | | |
|----|--|------|---|------|
| QY | | 61 | AAGACCGCGTGTGAGCACTCACCCTCTACTATCCAGTAGACGGACTTTGCCAAATGGC | 120 |
| Db | | 3768 | -TGATCCTCATCGGTGGTGAGGCCAATCCCACCTCCACTCACCGCACTTTTCGGTTGGC | 3826 |
| QY | | 121 | AGAAG-----GACCAAATTATAGAGCAGGAGAAGCAACTCAACTACTGTGAAGAACAAC | 174 |
| Db | | 3827 | AAAAGCAGGTGCRAACAGGAGACCGAACATAGCGGGCACTCGAATACTGGGTCAACGAGC | 3886 |
| QY | | 175 | TCAAAGACTCTTCCCCAGCAAAAGATCCGACCGAGCTTTGCGCCCTCGCACTTCTGTCTG | 234 |
| Db | | 3887 | TCGACAGACAGCTCGGCGCGGAATTCTTAACCGGACTTCCCCCGACCCAACATACTGTCCG | 3946 |
| QY | | 235 | GAGACGGAGGTTGCGGTACATGTTACATCGACGGCGAGCTCTACCAAGTCCCTTCGAGCCT | 294 |
| Db | | 3947 | GTGAAGCAGGTTCCGTCGCCAGTACGACATCGAAGCGGAACCTGTATGAAGAGCTCCAAGAAT | 4006 |
| QY | | 295 | TCTGCAACGAAACAACAACAGACCTCTTTCTGTCGTTCTTCTAGCTGGCTTCGCTGCCGCTC | 354 |
| Db | | 4007 | TCGTGAAGTAGAGCAAAATGACGCCCTTTCGCCGTGTGTGAGGGGCTTCGCGCGGACCC | 4066 |
| QY | | 355 | ATTATCGCTCACAGCGTGTGAAGACGCTGTCTATGGTACACCAATTTGCCAATCGCAACC | 414 |
| Db | | 4067 | ATTATCGCTCACCGGCGCGAAGACTCGATCATCGCAGCCCATCGCGAACCAGCAACC | 4126 |
| QY | | 415 | GACCTGAACCTGGAGGATATCATCGCGTGTCTTCTCAATAGCGAGTGTATGCGAATCAACA | 474 |
| Db | | 4127 | GCCAGGAGCTTGAAACATGATCGGCTTCTTCGTCACACCCAATGCATSCGAATCACGG | 4186 |
| QY | | 475 | TAGATCATCAGGATACGTTTGGGACTTTGATCAACCAAGTCAAGGCTACGACGACGACG | 534 |
| Db | | 4187 | TCGACGGCGACGACACTTTTGAAGCGTGTGTGCGGCAAGTTCGGAGCACCGCGACGCGG | 4246 |
| QY | | 535 | CATTGAGACGAGGATATTCCTCTTCAGCGCGTTGTATCAGCACTACAGCCTTGGATCCA | 594 |
| Db | | 4247 | CATTGAGACCAAGAGCGTCCCTTTGAGCGCGTCTGAGCGGCACTCCTTCCACGCTCGA | 4306 |
| QY | | 595 | GAGATCTGTCAAGCACACCTCTCGCAACTCATTTTTGCAGTGCACCTACAGAAGSACC | 654 |
| Db | | 4307 | GAGACCTATCCGAACCACTAGACAGCTACCTTCGCTCTTCATCTCACAGGACC | 4366 |
| QY | | 655 | TTGGAAGATTCAGTTCCAGGGTCTCGAGTCGTAAGTTCGCTAGCAAGCGTACACTC | 714 |
| Db | | 4367 | TCGGCAAGTTCGAGCTGGAGGGTCTCGTAGCGAACCCTGCTCGAACAAGGTATACCCA | 4426 |
| QY | | 715 | GATTTGACATGGAGTTCATCTGTTTCAAGAAACCGACAGCCTTAAAGGTAGCGTCAACT | 774 |
| Db | | 4427 | GGTTCGACGTGGAGTTTCACTCTGTCCAAGAACCGGAAGACTTAAGCGGTAAACGTGGCAT | 4486 |
| QY | | 775 | TTCCGATGAGCTGTTCAAAATGSAGACTCTTGAAATGTCGTCAGAGTATTTCTTTCAGA | 834 |
| Db | | 4487 | TTCCGGCAGATCTATTTCAGCCGTGAGACCAATTAGCAATGTAGTCGCCATATTTTCCAAA | 4546 |
| QY | | 835 | TTCTGAGAAACGGGCTTCAAAGTTTCGCGACACGAGTCTCAATACTTCTTTTCACATGATG | 894 |
| Db | | 4547 | TCCTGGCACAAGCATTCGCCACGCTTCGACTCCAATCGCTGTCTCTCCGCTTTACCGATG | 4606 |
| QY | | 895 | GCATTGTGACTTTGAAAAATTTGGATGTTCTCAACGTCAAACATGTGCGACTATPCCCGGAG | 954 |
| Db | | 4607 | GGTTAGCGGACCTTCGTGCCATGGGCTTGCTTCAGATCGAAGAGGCAAGTAATACC CGGG | 4666 |
| QY | | 955 | AATCGAGCTGGCTGATGCTTCCAGACCAAGTCTCTGCTTACC | 1000 |
| Db | | 4667 | AGTCTGAGCTGCTCGAGCTGTTCCGAAGCAGAGTGGCGCTCAACC | 4712 |

RESULTS

RESULT 3
AAQ54386

AAQ54386
ID AAQ54386 standard; DNA; 46899 BP.

XX
AC
AAQ54386;

XX

DT 08-JUL-1994 (first entry)

| | |
|----|--|
| XX | T. niveum Cyclosporin synthetase gene. |
| DE | |
| XX | Enzyme: cyclosporin; synthetase-like activity; Tolypocladium niveum; |
| KW | T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase; ss. |
| XX | |
| XX | Tolypocladium niveum. |
| XX | |
| PH | Key Location/Qualifiers |
| FT | CDS 885..46730 |
| FT | /*tag= a |
| FT | /product= Cyclosporin synthetase |
| FT | 40239..43129 |
| FT | /*tag= b |
| FT | /note= "SalI restriction fragment, preferred |
| FT | fragment, Claim 4" |
| FT | |
| FT | misc_feature 37781..40244 |
| FT | /*tag= d |
| FT | /note= "SalI restriction fragment, preferred |
| FT | fragment, Claim 5" |
| XX | |
| XX | EP578616-A. |
| PN | |
| XX | 12-JAN-1994. |
| PD | |
| XX | |
| XX | 05-JUL-1993; 93EP-0810474. |
| PF | |
| XX | |
| XX | 09-JUL-1992; 92AT-0001403. |
| PR | 08-MAR-1993; 93AT-0000437. |
| PR | 29-APR-1993; 93CH-0001310. |
| PR | 04-MAY-1993; 93CH-0001375. |
| XX | |
| XX | (SANO) SANDOZ LTD. |
| XX | (SANO) SANDOZ PATENT GMBH. |
| PA | (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH. |
| PA | |
| XX | Leitner E, Schneider E, Schoergendorfer K, Weber G; |
| PI | |
| XX | |
| XX | WPI: 1994-010432/02. |
| DR | P-PSDB; AAR44929. |
| DR | |
| XX | |
| PT | Isolated DNA sequence - which codes for enzyme having cyclosporin |
| PT | synthetase like activity |
| PT | |
| XX | |
| XX | Claim 6; Page 17-41; 93pp; English. |
| PS | |
| XX | |
| CC | This sequence encodes an enzyme which has cyclosporin synthetase- |
| CC | like activity. This sequence was isolated from Tolypocladium niveum |
| CC | (formerly known as T. inflatum GAMS). The enzyme encoded by this |
| CC | sequence catalyses the peptide biosynthesis of cyclosporins and |
| CC | structurally related molecules. This sequence may be used for the |
| CC | production of cyclosporin by transforming a vector containing this |
| CC | sequence in to a recombinant host. This allows effective production |
| CC | of antibiotic cyclosporin or its derivatives. |
| XX | |
| XX | Sequence 46899 BP; 10651 A; 13513 C; 12509 G; 10226 T; 0 other; |
| SQ | |

DT 08-JUL-1994 (first entry)

| | | |
|-------|--|-------|
| 178 | AGAGCTTCTCCGACGAAAGATCCCGACGAGCTTTGCCGCGCTTCACTTCTGTCTGGAG | 237 |
| | | |
| 30403 | CTCAGAGTACTCCGGTGAGCTTCTGAGCCAGCTTTCCCGCCACAGTTCTGAGTGGA | 30462 |
| | | |
| 238 | ACGCAGGTTGCGGTACATGTTACCATCGAGCGGCGAGCTCTACCACTCCCTTCGAGCCTTCT | 297 |
| | | |
| 30463 | AGCTGGTGTATCCCGGTCACCATGAGGGGCGGTCTACGAGAAGCTTCTGAAGTTCT | 30522 |
| | | |
| 298 | GCAAGAACACACACGAGCCTCTTTCGTCGTCTTCTTCTAGCTGCGTTCGGTCCGCTCATTT | 357 |
| | | |
| 30523 | CCAAGGAGCGCCAGGTAACCTGTCTCGGTCTATTAAACAGGTTCCGGGCCACACT | 30582 |
| | | |
| 358 | ATCGTCTCACAGCTGTTGAAGACGCTGTCAATTGGTACACCAATTGCGAATCGCAACCGAC | 417 |
| | | |
| 30583 | TTTCGTCTCACTGGTCAGAGGATGCTACGATCGGTACCCCAATTGCAAAATCGCAACGGC | 30642 |
| | | |
| 418 | CTGAAGTCGAGGATATCATCGGCTGCTTTGTCTAATACGACGTGTATGGATCAACATAG | 477 |
| | | |
| 30643 | CAGAAGTCAGCATATCATTTGATTTCTGCTCAACCCCAATGATGGTCTTCTCCTCG | 30702 |
| | | |
| 478 | ATCATCACGATACCTTTGGGACTTTGATCAACCAGTCAAGGCTACGACGACGACGAT | 537 |
| | | |
| 30703 | ATACCGGCAGCACATTCGAAATCCCTAGTCCAGCATGTTCCGTCGCTGCTACAGATGCT | 30762 |
| | | |
| 538 | TCGAGAACGAGATATTCGGTTGAGCGGTTGATCAGCACTACAGCCTGGATCCAGAG | 597 |
| | | |
| 30763 | ATTCCATCAGGATATTCCTTCGAACGGATGCTCTCGGCACTTCTCCTCGGCTCGAGAG | 30822 |
| | | |
| 598 | ATCTGTCAAGCACACCTCTCGCACAACTCATTTTTGCAAGTCACTCAAGAACGACCTTG | 657 |
| | | |
| 30823 | ATGCCTCAGAAAGCCCACTAATCCAGCTTATGTTTGCTTGCATCAGCCACATCTCG | 30882 |
| | | |
| 658 | GAAGATTCAAGTTCCAGGCTCTCGAGTCCGTACCTGTGCTGCTAGCAAGCGTACACTCGAT | 717 |
| | | |
| 30883 | GGAACATTACTCTGAAGGACTCTGAGCATGAGCGCTGCCAACAAAGCGTCCCAACAGGTT | 30942 |
| | | |
| 718 | TTGACATGAGTTCCATCTGTTTCAAGAAACCGACAGCCTTAAAGGTAGCGTCAACTTG | 777 |
| | | |
| 30943 | TCGACATGAGTTCCACCTGTTCGAAGAGCCTACAGCTGAGTGGTTTCAATCTCTTG | 31002 |
| | | |
| 778 | CCGATGAGCTGTTCAAAATGGAGACTGTTGAAAATGTGCTCAGAGTATTTCTTGAGATTC | 837 |
| | | |
| 31003 | CCGATGAGCTTTCAGAGCTGAAACAAATCAACAGCGCTGCTGACTGTGTTCCAGGAGATAC | 31062 |
| | | |
| 838 | TCGAAACGGGCTTCAAAGTTCCGGCACACAGTCTCAATACTTCCCTTTGACTGATGGCA | 897 |
| | | |
| 31063 | TCGAGCGGGGCTTCGACCAACCCCAAGTCTCCATTTCTACTATGCCCTGACTGATGGGT | 31122 |
| | | |
| 898 | TTGTGACTCTTGAAAAATTTGGATGTTCTCAAGTCAAAACATGTGCGACTATTCGCCGGAAT | 957 |
| | | |
| 31123 | TGATTGATCTCGAAGAACTGGGCTTGCTGGAATCGAGACGACAACTTCCCTCGGACT | 31182 |
| | | |
| 958 | CGAGCTTGGCTGATGCTTCCAGACCCCAAGTCTCTGCTTACCC | 1000 |
| | | |
| 31183 | ACTCGGTTGTTCGACGCTTCTCCGACAGCAGGTTGGTGGCAATCC | 31225 |
| | | |

| | |
|------------|--|
| RESULT 4 | |
| AAQ78276/C | |
| ID | AAQ78276 standard; DNA; 3973 BP. |
| XX | |
| XX | |
| AC | AAQ78276; |
| XX | |
| XX | |
| DT | 06-JUN-1995 (first entry) |
| XX | |
| XX | |
| DE | T. niveum alanine-racemase. |
| XX | |
| XX | |
| KW | Alanine-racemase; glyceraldehyde-3-phosphate-dehydrogenase |
| KW | cyclosporin; immunosuppressive; ss. |
| XX | |
| OS | Tolypocladium niveum ATCC 34921. |
| XX | |
| PN | W09425606-A. |

[illegible]

QY 538 TCGAGACGAGGATATTCGGTTTGAGCGCGTTGTATACGCACTACAGCCTGGATCCA 594
 ID AAQ78281/c
 DB 58 ATGCCAACCGCGACGCTCCCTTTGAGCGCATCGTGTCTGCTGCTGCCCGGTCGA 2

RESULT 5
 AAQ78281/c
 ID AAQ78281 standard; DNA; 3973 BP.
 AC AAQ78281;
 XX
 DT 06-JUN-1995 (first entry)
 XX
 DE T. niveum alanine-racemase.
 XX
 KW Alanine-racemase; glyceraldehyde-3-phosphate-dehydrogenase;
 KW cyclosporin; immunosuppressive; ss.
 XX
 OS Tolypocladium niveum ATCC 34921.
 XX
 FH Key Location/Qualifiers
 FT 1953..3180
 CDS /tag= a
 FT intron 2035..2110
 FT /tag= b
 FT
 XX W09425606-A.
 XX
 PD 10-NOV-1994.
 XX
 PF 23-APR-1994; 94WO-EP01272.
 XX
 PR 23-APR-1993; 93DE-4312856.
 PR 04-MAY-1993; 93DE-4314610.
 PR 17-MAY-1993; 93DE-4316419.
 XX
 PA (SANO) SANDOZ LTD.
 PA (SANO) SANDOZ PATENT GMBH.
 PA (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
 PI Kocher HP, Schneider-Scherzer E, Schoergendorfer K;
 PI Weber G;
 XX
 DR WPI; 1994-358281/44.
 DR P-PSDB; AAR65964.
 XX
 PT New nucleic acid encoding eukaryotic alanine racemase - and
 PT related vectors, host cells and recombinant enzyme, useful for
 PT producing cyclosporin derivs. or increasing cyclosporin prodn.,
 PT also new glyceraldehyde-3-phosphate dehydrogenase gene.
 XX
 PS Disclosure; Fig.20; 82pp; German.
 XX
 SS A genomic DNA library of T. niveum ATCC 34921 was screened with probes
 CC based on the isolated alanine-racemase (AR) enzyme. Lambda clone RAC4
 CC contained a 1.1 kb PstI fragment, a 1.9 kb EcoRI-SalI fragment and a
 CC 650 HindIII-PstI fragment, which were subcloned in plasmid vectors
 CC and combined to give the sequence of the AR DNA (see also AAQ78276).
 XX
 SQ Sequence 3973 BP; 917 A; 1035 C; 1104 G; 917 T; 0 other;

Query Match 16.8%; Score 168.6; DB 15; Length 3973;
 Best Local Similarity 56.4%; Pred. NO. 1.2e-43;
 Matches 337; Conservative 0; Mismatches 254; Indels 6; Gaps 1;

QY 4 TTGATGCTCTGGACCGGATCTCAATCAGCTTACTCAGCTGGCTCAAGGACTCAAAAG 63
 DB 598 TCGACATATTTCAGCAGGAGCTGGAGAACTCTACACGGCCGCTCGCAGGGGAAATCGA 539
 QY 64 ACCCGCTGTGACGACTCACTCTCTACCTATCATCAGTACAGCGACTTTGCAAAATGGCAGA 123
 DB 538 TTTCGGCTGTCCCTTGGGTCCCAATTCCCATTCATACCTGACTTGACGACTTGGCAGA 479

QY 124 -----AGGACCAATTCATAGACGAGGAAGCAACTCAACTACTGGAAGAAGCAACTCA 177
 DB 478 ACCAGGACGAGCAGGTGCTGCTGAGCAGGAAGGACGCTCGATATCGATCGAGCAGCTCG 419
 QY 178 AAGACTCTTCCCGCAGCAAGATCCCGACCGACTTTCCCGCCCTCGACTTCTGTCTGGAG 237
 DB 418 ATAACAACACACCGCGCAGCTCCTCACAGAGCTTCCCGCGCAGCATATCCCATCTTGGG 359
 QY 238 ACGCAGTTTGCCTACATGTTACCATCGACGGCGAGCTCTACCACTCCCTTCGAGCCTTCT 297
 DB 358 AAACCTGGCAAGATCTCCTTCCAGATCGATGGATCGGTACACAAGAACTCTCTGGCCCTTCT 299
 QY 298 GCAACGAAACACACGACGACCTCTTTCGTCGTTCTTCTAGCTGCGTTCCGTCGCGTCAAT 357
 DB 298 GCCGCTCCCGCAGCAAGTAACCGCTACGCCGTGCTGCTGCGAGGTTTGGCTGGCGCACT 239
 QY 358 ATCGTCTCACAGCTGTTGAAGAGCGCTGTCTATGTTGGTACACCAATTTGGGAATCGCAACCGC 417
 DB 238 TTCGCTCACTGGAGCCGAGGATGCAACCATCGGAGCGCCGTTGCCAACCGCGACCGC 179
 QY 418 CTGAACCTGGAGGATATCATCGGCTGCTTGTCAATACGCGAGTGTATCGGAATCAACATAG 477
 DB 178 CGGAGCTGGAGAACATGGTGGCTCCCTTGGCCACTCTGCAGTGCATGCGAGTGTGCTCG 119
 QY 478 ATCATCAGATACCTTTGGGACTTTTGATCAACCAAGTCAAGGCTACGACGACAGCAGCAT 537
 DB 118 ACGAGGACGACACCTTCGAGTCGGTCTGCGGAGATCATGTCCGTCATGACAGAGGCAC 59
 QY 538 TCGAGAACGAGGATATTCGTTTGAGCGGCTGTATCAGCACTACAGCCTGGATCCA 594
 DB 58 ATGCCAACCGCGACGCTCCCTTTGAGCGCATCGTGTCTGCGTTGCTGCCCGGTCGA 2

RESULT 6
 AA11992
 ID AA11992 standard; DNA; 37856 BP.
 XX
 AC AA11992;
 XX
 DT 07-AUG-2000 (first entry)
 XX
 DE S. cellulosum DNA encoding polyketide and heteropolyketide enzymes.
 XX
 KW Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;
 KW epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;
 KW plant-protection; ds.
 XX
 OS Sorangium cellulosum.
 XX
 FH Key Location/Qualifiers
 FT CDS complement (3398..6100)
 FT /tag= a
 FT /product= "ORF1-trRNA synthetase"
 FT /note= "gtg start codon"
 FT CDS complement (6374..7111)
 FT /tag= b
 FT /product= "ORF2-monoxygenase"
 FT CDS complement (8433..9550)
 FT /tag= c
 FT /product= "ORF3-aminotransferase"
 FT /note= "AGT start codon given in the specification"
 FT CDS 9855..11393
 FT /tag= d
 FT /product= "ORF4- tyrosine/DOPA-Decarboxylase"
 FT /note= "GTG start codon"
 FT CDS 12212..13658
 FT /tag= e
 FT /product= "ORF5-3-oxoacyl-ACP-reductase"
 FT /note= "ACC start codon"
 FT CDS 15374..19984
 FT /tag= f
 FT /product= "ORF6-polyketide synthase"
 FT CDS 20003..27889

```

FT FT      /*tag= 9 "ORF7-peptide synthetase"
FT FT      28251..29400
FT FT      /*tag= h
FT FT      /product= "ORF8-transpeptidase"
FT FT      complement (30040..31720)
FT FT      /*tag= i
FT FT      /product= "ORF9-regulation element"
FT FT      /note= "CGC stop codon"
FT FT      31982..32932
FT FT      /*tag= j
FT FT      /product= "ORF10-transcription regulator"
FT FT      33128..33613
FT FT      /*tag= k
FT FT      /product= "ORF11-regulation element"
FT FT      /note= "GTG start codon"
FT FT      33661..34077
FT FT      /*tag= l
FT FT      /product= "ORF12-regulation element"
FT FT      complement (35255..35616)
FT FT      /*tag= m
FT FT      /product= "ORF13-transcription regulator"
FT FT      complement (35730..36242)
FT FT      /*tag= n
FT FT      /product= "ORF14-transcription regulator"
FT FT      /note= "GTG start codon"
XX XX
XX XX      DE19846493-AL.
XX XX      13-APR-2000.
XX XX      09-OCT-1998; 98DE-1046493.
XX XX      09-OCT-1998; 98DE-1046493.
XX XX      (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX XX      Beyer S, Mueller R;
XX XX      WPI; 2000-294101/26.
XX XX
XX XX      DNA sequence coding for products involved in the biosynthesis of
XX XX      polypeptide or heteropolypeptide compounds, especially epothilone
XX XX
XX XX      Claim 3; Page 20-33; 36pp; German.
XX XX
XX XX      This invention describes a novel DNA sequence (I) whose expression
XX XX      products effect or are involved in the enzymatic biosynthesis,
XX XX      mutasynthesis or partial synthesis of polypeptide or heteropolypeptide
XX XX      compounds (II). (I) can be inserted into an expression vector and used
XX XX      to transform or transfect prokaryotic or eukaryotic cells with the aim
XX XX      of obtaining strains that produce large amounts of polypeptide or
XX XX      heteropolypeptide compounds, especially epothilones, which have cytotoxic
XX XX      and/or immunosuppressant and antibiotic and antifungal activities and
XX XX      are useful as plant-protection agents. This sequence represents the DNA
XX XX      sequence isolated from Sorangium cellulosum which is described in the
XX XX      method of the invention.
XX XX
XX XX      Query Match      8.6%; Score 85.6; DB 21; Length 37856;
XX XX      Best Local Similarity 50.2%; Pred. No. 4.6e-16;
XX XX      Matches 270; Conservative 0; Mismatches 259; Indels 9; Gaps 2;
XX XX
XX XX      61 AGAGCCCGCTGTGCAGCACTCCTCTACCTATCCAGTACAGCGACTTTGCAAAATGGC 120
XX XX      20535 AGGCGCGCGTGTCCGCGCTCCGAGTCCGATTCATACGCGGACTTCGCGTGTGGC 20594
XX XX
XX XX      121 AGAAGGA-----CCAAATTCATAGCAGGAGAGCAACTCACTTGGAGAGCAAC 174
XX XX      20595 AGCGGGAGCTGTCCAGGGGGAAGTCTTGAATCGCACTCGGCTACTGGAGAGCAACC 20654
XX XX
XX XX      175 TCAAGACTCTTCC---CCAGCAAGATCCGACCGACTTTGCCCGCCCTGCCTCTGT 231
XX XX

```

Db 20655 TCCGCGCGCCCCCACGCTGCTGGAGCTTCGGATGGACCGCCCGCGCGCGCAGA 20714
 QY 232 CTGGAGACGCGAGGTTGGGTACATGTTACCATCGACGCGGAGCTCTACCAAGTCCCTTCGAG 291
 Db 20715 CGTTCGGGGCTCCAGCGCGGCTTCGAGCTCCCACTCTCCTCTGCAACAGCGGTGAGG 20774
 QY 292 CCTTCTGCAACGAACAACACACGACCTCTTTTCGCTGTTCTTCTTCTAGCTGCGTTCGTCGCG 351
 Db 20775 CGCTCAGCGCGCAGGAAGCGCGACCCCTTCATGACGCTCTGACGCGGTTACGCGTGC 20834
 QY 352 CTCATTATGCTCTCACAGCTGTTGAAGAGCGCTGTCATTGGTACACCAATTCGGAATCGCA 411
 Db 20835 TGCTCTCGCGTTATGCGCGCAGAGCATGCTGGTGGCACGCCCATCGCAATCGCA 20894
 QY 412 ACCGACCTGAACCTGGAGGATATCATCGGCTGCTTTGTCAATACGCACTGTATGCAATCA 471
 Db 20895 CCCGAGCAGCTGGAGGGGCTGATCGGCTTCTTCGTCACATCTGCTGGCGTGGCATCG 20954
 QY 472 ACATAGATCATCAGATACCTTTGGGACTTTTGATCAACCAAGTCAAGGCTACGACGACAG 531
 Db 20955 ACCTCGGGGCGCACCAGCTTCGCGAGCTGCTCGGGCGGTGCGGGAGGTGAGCTTGG 21014
 QY 532 CAGCATTCGAGACGAGGATATTCCTGTTGAGCGGCTTGTATCAGCACTACAGCGCTGG 589
 Db 21015 GCGCCTACGCGCACGAGGACCTGCCCTTCGAACGGCTGGTGGAGGAGCTGTACACAGG 21072

RESULT 7
 AAS51470
 ID AAS51470 standard; DNA; 13029 BP.
 XX
 AC AAS51470;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Pseudomonas aeruginosa DNA for cellular proliferation protein #55.
 XX
 KW Antisense: ds; prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR P-PSDB; AAU33611.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 XX
 PS Claim 27; Seq ID No 4052; 51lpp; English.
 CC
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential

PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DRFZ-) DRFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
PI Fraser CM, Tuemmler B, Hohelsel J, Duesterhoeft A, Hilbert H;
PI Timmis KN, Moore E, Straetz M, Heim S, Golyshin P;
XX WPI; 2001-159715/16.
XX
XX New DNA encoding a bacterial lipopeptide synthase, useful e.g. for
PT producing antimicrobial agents, surfactants with increased
PT bioavailability in microbial depollution methods, and transgenic plants
PT
XX
PS Claim 3a; Page 28-30; 33pp; German.
XX
XX This invention describes a novel DNA sequence (I) that expresses products
CC having the biological function of lipopeptide synthase. The invention
CC also describes (1) recombinant expression vectors containing (1); (2)
CC prokaryotic and eukaryotic cells transformed or transfected with (1) or
CC the vector of (2); (3) production of lipopeptide synthases by culturing
CC cells of (2); (4) expression products (II) of (1), and synthetic proteins
CC or peptides with the same sequences; (5) mono- or poly-clonal antibodies
CC (Ab) specific for (II); (6) hybridoma cells that produce monoclonal Ab;
CC and (7) transgenic plants that contain cells of (2). (1), and their
CC fragments, are useful for expression of recombinant lipopeptide
CC synthases, and as probes and primers for detection, isolation and
CC amplification of full-length cDNA sequences. (1) are used to produce
CC transgenic plants. Lipopeptide synthases are useful for production of
CC antiviral, antibacterial or antifungal lipopeptides, particularly useful
CC for treatment of plants, and for production of lipopeptide surfactants
CC used for increasing bioavailability in microbial depollution processes
CC (acting as emulsifiers) and in tertiary crude oil recovery. Individual
CC subunits from different lipopeptide synthases may be combined to allow
CC synthesis of many different biologically active substances.
XX
SQ Sequence 5451 BP; 937 A; 1766 C; 1807 G; 941 T; 0 other;

Query Match 7.08; Score 70.4; DB 22; Length 5451;
Best Local Similarity 48.24; Pred. No. 1.3e-11;
Matches 265; Conservative 0; Mismatches 276; Indels 9; Gaps 2;

Qy 48 CTCAGGACTCAAAAGACCGCTGTACGACTACTCTCTACCTATTCACGACGAC 107
Db 874 CACCATCCCGCAGGGCGAGCGCCGTGATTTCCGCGCTTTCGGTGCATACCGCGAC 933
Qy 108 TTTCGCAAAATGGCAGAGACCAATTCATAGACGAGGAGCAAA-----CTCAACTAC 161
Db 934 TATGCCAGCTGGCAACGCGAGCGCTGAGCGAAGGCCAGATGCAAGCCAGCTCGGGCTAC 993
Qy 162 TGGAGAGCACTCAAGACTCTTCCCGAG---CAAAGATCCCGACCGACTTGGCCCGC 218
Db 994 TGGCAGCGGCACCTGGGAAGATGATTTTCGAGTGTCTGACGTACCTGCGCGACCGCCACGG 1053
Qy 219 CCTGCACCTTGTCTGGAGACGAGGTTCGTATCATGTTACCATCGACGCGAGCTCTAC 278
Db 1054 CCCGAGCTGCAAGCTATCGGGCGGACGTCTGGACGTGCGCTGCGCGCGAGTTGACG 1113
Qy 279 CAGTCCCTTCGAGCCTTCTGCAACGAACACACAGCACTCTTTTCGTCGTTCTTCTAGCT 338
Db 1114 GCGGACCTGCGGCGCTGGCAGTGCAGTGCAGCCAGCGTGTTCACAGCTCTCTCTAGCC 1173
Qy 339 CGGTTCCGTCGCTCATTTATGCTCTACAGCTGTTGAAGACGCTGTCTATTGGTACACCA 398
Db 1174 TCGTTCGGCTGTGCTGCGCGCTACAGCGGCAACGAGAAGATCAACATCGCGTGCCTC 1233
Qy 399 ATTGCGAATCGCAACCGACCTGAACCTGGAGGATATCATCGGCTGCTTGTCAATACGACG 458
Db 1234 ATGACCAACCGCAATCCCTTGGAGCTGAGAGGTTGATCGGGTCTTCTGCTCAACACCGTG 1293
Qy 459 TGTATGCAATCAACATGATATCATGATACCTTTTGGGACTTTGATCAACACCAAGTCAAG 518
Db 1294 GTGCTGCGGCTGAGTGTGGACTACAGCGCTCAGCTTTCGAGCAGTGTGCTGGCCCATACCAA 1353

Qy 519 GCTACGACGACGACGACGACGACGATTCGTTTGTAGCCGCTTGTATCAGCA 578
Db 1354 GAATATCAGCTGCGAGGCCAGGCCCAACAGGACCTGCCATTCGACGCTCTGTCGAGCG 1413
Qy 579 CTACGACGCTG 588
Db 1414 CTGACGCCAG 1423
RESULT 12
AAAL4651
ID AAAL4651 standard; DNA; 77536 BP.
XX
XX AAAL4651;
XX
XX 08-AUG-2000 (first entry)
XX Nucleotide sequence of the FK-520 biosynthetic gene cluster.
DE
XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW FK-506 binding protein; polyketide compound; transplant rejection;
KW graft-versus-host disease; uveitis; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
XX Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS complement (412..1836)
FT FT /*tag= a
FT FT /*note= "fkBW gene"
FT FT complement (2020..3579)
FT FT /*tag= b
FT FT /*note= "fkBV gene"
FT FT 3969..4496
FT FT /*tag= c
FT FT /*note= "fkBR2 gene"
FT FT complement (4595..5488)
FT FT /*tag= d
FT FT /*note= "fkBR1 gene"
FT FT 5601..6818
FT FT /*tag= e
FT FT /*note= "fkBE gene"
FT FT 6808..8052
FT FT /*tag= f
FT FT /*note= "fkBF gene"
FT FT 8156..8824
FT FT /*tag= g
FT FT /*note= "fkBG gene"
FT FT complement (9122..9883)
FT FT /*tag= h
FT FT /*note= "fkBH gene"
FT FT complement (9894..10994)
FT FT /*tag= i
FT FT /*note= "fkBI gene"
FT FT complement (10987..11247)
FT FT /*tag= j
FT FT /*note= "fkBJ gene"
FT FT complement (11244..12092)
FT FT /*tag= k
FT FT /*note= "fkBK gene"
FT FT complement (12113..13150)
FT FT /*tag= l
FT FT /*note= "fkBL gene"
FT FT complement (13212..23988)
FT FT /*tag= m
FT FT /*note= "fkBC gene"
FT FT complement (13452..13662)
FT misc_feature

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 19:04:47 ; Search time 57 Seconds
(without alignments)
5385.674 Million cell updates/sec

Title: US-09-482-788-1_COPY_4000_5000

Perfect score: 1001

Sequence: 1 caattgatgtctgcagcgc.....cccaagtctgtgtacc 1001

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCFUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|---------|-------|-------------------|
| 1 | 339.4 | 33.9 | 46899 | 1 | US-08-471-119A-1 |
| 2 | 62.8 | 6.3 | 178 | 1 | US-08-471-119A-3 |
| 3 | 60.4 | 6.0 | 474 | 2 | US-08-403-852D-14 |
| 4 | 60.4 | 6.0 | 474 | 3 | US-08-510-646B-14 |
| 5 | 60.4 | 6.0 | 474 | 4 | US-09-231-818-14 |
| 6 | 60.4 | 6.0 | 4403765 | 4 | US-09-103-840A-2 |
| 7 | 60.4 | 6.0 | 4411529 | 4 | US-09-103-840A-1 |
| 8 | 53.6 | 5.4 | 2219 | 3 | US-08-510-646B-17 |
| 9 | 47.2 | 4.7 | 4403765 | 4 | US-09-103-840A-2 |
| 10 | 47.2 | 4.7 | 4411529 | 4 | US-09-103-840A-1 |
| 11 | 41.2 | 4.1 | 11601 | 2 | US-08-222-617A-3 |
| 12 | 41.2 | 4.1 | 11601 | 2 | US-08-222-617A-24 |
| 13 | 33.3 | 3.3 | 330 | 2 | US-08-743-200-5 |
| 14 | 33.3 | 3.3 | 8532 | 1 | US-08-452-655B-1 |
| 15 | 33.3 | 3.3 | 8532 | 3 | US-08-450-582-1 |
| 16 | 33.3 | 3.3 | 9606 | 1 | US-07-741-940-1 |
| 17 | 33.3 | 3.3 | 9606 | 1 | US-08-289-548A-1 |
| 18 | 33.3 | 3.3 | 9606 | 1 | US-08-452-654-1 |
| 19 | 33.3 | 3.3 | 9606 | 2 | US-08-370-235A-1 |
| 20 | 33.3 | 3.3 | 9606 | 4 | US-08-449-731-1 |
| 21 | 32.8 | 3.3 | 6827 | 1 | US-08-222-616-17 |
| 22 | 32.8 | 3.3 | 6827 | 4 | US-08-446-648-17 |
| 23 | 32.8 | 3.3 | 6827 | 5 | PCT-US95-04228-17 |
| 24 | 32.6 | 3.3 | 7218 | 1 | US-08-232-463-14 |
| 25 | 32.6 | 3.3 | 26385 | 4 | US-08-961-527-3 |
| 26 | 31.8 | 3.2 | 5365 | 4 | US-08-961-527-77 |
| 27 | 31.4 | 3.1 | 807 | 1 | US-07-706-691G-97 |

| | | | | | |
|----|------|-----|------|---|-------------------|
| 28 | 31.4 | 3.1 | 807 | 1 | US-08-254-021-97 |
| 29 | 31.4 | 3.1 | 807 | 2 | US-08-618-446-97 |
| 30 | 31.4 | 3.1 | 807 | 3 | US-08-980-135-97 |
| 31 | 31.4 | 3.1 | 807 | 4 | US-09-585-798-97 |
| 32 | 31.4 | 3.1 | 1395 | 4 | US-09-586-719-3 |
| 33 | 31.4 | 3.1 | 3601 | 4 | US-08-334-179A-1 |
| 34 | 31.4 | 3.1 | 3603 | 4 | US-08-462-467B-1 |
| 35 | 31.4 | 3.1 | 4776 | 2 | US-08-852-401-1 |
| 36 | 31.2 | 3.1 | 4195 | 1 | US-08-340-011-1 |
| 37 | 31.2 | 3.1 | 4195 | 3 | US-08-901-710-1 |
| 38 | 31.2 | 3.1 | 4416 | 3 | US-08-795-430-1 |
| 39 | 31.2 | 3.1 | 4416 | 4 | US-09-353-700-1 |
| 40 | 31.2 | 3.1 | 4416 | 4 | US-08-601-132-36 |
| 41 | 31.2 | 3.1 | 4425 | 1 | US-08-222-616-31 |
| 42 | 31.2 | 3.1 | 4425 | 4 | US-08-446-648-31 |
| 43 | 31.2 | 3.1 | 4425 | 5 | PCT-US95-04228-31 |
| 44 | 31.2 | 3.1 | 4795 | 1 | US-08-340-011-3 |
| 45 | 31.2 | 3.1 | 4795 | 3 | US-08-901-710-3 |

ALIGNMENTS

RESULT 1

US-08-471-119A-1
; Sequence 1, Application US/08471119A
; Patent No. 5827706

GENERAL INFORMATION:

APPLICANT: Leitner, Ernst
APPLICANT: Schneider, Elisabeth
APPLICANT: Schoergendorfer, Kurt
APPLICANT: Weber, Gerhard
TITLE OF INVENTION: Cyclosporin Synthetase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. 5827706artis Corporation
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,119A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kassenoff, Melvyn
REGISTRATION NUMBER: 26,389
REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
TELEPHONE: 201 503 8474
TELEFAX: 201 503 8807

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 46899 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Tolypocladium niveum
STRAIN: ATCC 34921

US-08-471-119A-1

Query Match 33.9%; Score 339.4; DB 1; Length 46899;
Best Local Similarity 60.1%; Pred. No. 2.2e-101;

Matches 603; Conservative 0; Mismatches 391; Indels 9; Gaps 2;

QY 4 TTGATGCTTTGGACGCGATCTCAATCAGCTCTACTCAGCTGGCTCAAGGACTCAAAAG 63
 Db 30226 TTGATATCTTGAGACGAGCTCGGTCAATCTACTCAGCGCTTTACGTGGC---AGGG 30282

QY 64 ACCGGCTGTGACGACCTCTCTCTACTATCAGTACAGCGGACTTTGCAAAATGGCAGA 123
 Db 30283 ATCCGTTATCTCAGGTCAAGCCCTCCCAATACTATCTGACTTTGGCGCTTGGCAGA 30342

QY 124 AGGA-----CCAAATCATAGACGAGGAGCAACTCAACTACTGTGAAAGCAACTCA 177
 Db 30343 AGGAGCTGCCAAGTTCGCGACATGAGAGGAGCTCGGTACTGGGAGAACAGTTAG 30402

QY 178 AAGACTTCTCCCGACAAAGATCCGACCGACTTTGCCCGCCCTGCGACTTCTCTGTGGAG 237
 Db 30403 CTGACAGTACTCCGGTGAGCTTCTGACCGACTTTCCCGCCACACGTTCTCTGATGGGA 30462

QY 238 ACCGAGTTGCTTACATGTTTACCATCGACGGCGAGCTCTACAGTCCCTTCGAGCGCTTCT 297
 Db 30463 AGCTGCTGTCATCCCGGTACCATTTGAGGGGCGGTCTACGAGAAGCTTCTGAAGTCT 30522

QY 298 GCAACGAACACACAGGACCTCTTTCGCTGCTTCTTAGCTGCTGCTTCGCTCGCTCATT 357
 Db 30523 CCAAGGAGCCAGGTAACCTCTGTTCTCGGTCTATTACAGGCTTCGCGGCCACACACT 30582

QY 358 ATCGTCTCAGCTGTTGAAGAGCGTGTCTANTGGTACACCAATTCGGAATCGCAACCGAC 417
 Db 30583 TTGCTCTCACTGTGACGAGGATCTACGATCGGTACCCCAATTTGCAAAATCGCAACCGGC 30642

QY 418 CTCAACTGGAGGATATCATCGGCTGCTTCTCAATACGCGAGTGTATCGGAATCAACATAG 477
 Db 30643 CAGAACTCGAGCATATCATTTGGATTTCTGTCACACCCCAATGCGTCTTCTCTCTCG 30702

QY 478 ATCATCAGCATACCTTTGGGACTTTGTAACCAAGTCAAGGCTACGACGACGACGACAT 537
 Db 30703 ATACCGGACGACATTCGATCCCTAGTCCAGCATGTTCCGTCGCTGGCTACAGATGCT 30762

QY 538 TCAGAACGAGGATATTCGCTTTGAGCGGCTTGTATCAGCACTACAGCTGGATCGACGAG 597
 Db 30763 ATTCCAATCAGGATATTTCCCTTCGAACGAGATCTCTCGGCACTTCTCCCTGGCTCGAG 30822

QY 598 ATCTGTCAAGCACACCTCGCACAACTATTTTTCAGTGCAGTGCAGTGCAGGAGGACTTG 657
 Db 30823 ATGCTTCAAGACCCCACTAATCCAGCTTATGTTTGCCTTGCCTACAGCCAGATCTCG 30882

QY 658 GAAGATTCAAGTTCCAGGCTCTCGAGTCCGTACCTGTGCTAGCAAGGCTACACTCGAT 717
 Db 30883 GGAACATTACTCTCGAAGGACTCGAGCATGAGCGCTGCCAACAGCGTTCGCAACAGCT 30942

QY 718 TTGACATGAGTTCATCTGTTTCAAGAACCGACAGCCTTAAAGTAGGCTCAACTTTG 777
 Db 30943 TCACATGAGTTCACCTGTTTCAAGAGCCTAACAGGCTGAGTGTGTTCAATCTCTTTG 31002

QY 778 CCGATGAGCTGTTCAAAATGGAGACTGTTGAAATGCTGTCAGAGTATCTTTGAGATTC 837
 Db 31003 CCGATGAGCTTCTCCAGCTTGAACCAATCAACAGCGCTGAGTGTTCAGGAGATAC 31062

QY 838 TGAGAACGGGCTTCAAAAGTTCGCGACACCAAGTCTCAATACTTCTTTGACTGATGGCA 897
 Db 31063 TCGAGCGGCGCTTCGACCAACCCCAAGTCTCCATTTCTACTATGCCCCCTGACTGATGGT 31122

QY 898 TTGTGACTTTGAAATTTGGATGTTCTCAAGCTCAACATGTCGACTATCCCGAGAT 957
 Db 31123 TGATTGATTCGAGAAATGGGCTGCTGGAAATCGAGAGCAACTTCCCTCGGACT 31182

QY 958 CGAGCTTGGCTGATGCTTCCAGACCCCAAGTCTCTGCTTACCC 1000
 Db 31183 ACTCGTTCGACGCTTCTCCGACGAGGCTGCTGCCAATCC 31225

RESULT 2
 US-08-471-119A-3

: Sequence 3, Application US/08471119A
 : Patent No. 5827706
 : GENERAL INFORMATION:
 : APPLICANT: Leitner, Ernst
 : APPLICANT: Schneider, Elisabeth
 : APPLICANT: Schoerendorfer, Kurt
 : APPLICANT: Weber, Gerhard
 : TITLE OF INVENTION: Cyclosporin Synthetase
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: No. 5827706artis Corporation
 : STREET: 59 Route 10
 : CITY: East Hanover
 : STATE: New Jersey
 : COUNTRY: USA
 : ZIP: 07936
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/471,119A
 : FILING DATE: 06 JUN-1995
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Kassenhoff, Melvyn
 : REGISTRATION NUMBER: 26,389
 : REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 201 503 8474
 : TELEFAX: 201 503 8807
 : INFORMATION FOR SEQ ID NO: 3:
 : LENGTH: 178 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: unknown
 : MOLECULE TYPE: DNA (genomic)
 : HYPOTHETICAL: NO
 : ANTI-SENSE: NO
 : ORIGINAL SOURCE:
 : ORGANISM: Tolypocladium geodes
 : US-08-471-119A-3

Query Match 6.3%; Score 62.8; DB 1; Length 178;
 Best Local Similarity 61.7%; Pred. No. 3.5e-11;
 Matches 100; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 387 ATTGGTACACCAATTCGGAATCGCAATCGCAACCGACCTGGAAGATATCATCGGCTGCTTT 446
 Db 9 ATCGGCTCTCCAATTCGGAACAGAAATCGAGCAGAGCTTGAGGCGCTTATTGGCTGTTT 68

QY 447 GTCAATACCGAGTGTATGGAATCGCAATCGCAATCGCAATCGCAATCGCAATCGCAAT 506
 Db 69 GTGATACCTAGTGTATGAGACTGCCAGTTACCGATGAGATACATTTCGCCAATTGATT 128

QY 507 AACCAAGTCAAGGCTACGACGACGACGACGACGACGACGACGACGACGACGACGACG 548
 Db 129 GACTGTGTACGAGAGAGCTCAACCGAGGCTTTCGACCAACGACGACGACGACGACG 170

RESULT 3
 US-08-403-852D-14
 : Sequence 14, Application US/08403852D
 : Patent No. 5891695
 : GENERAL INFORMATION:
 : APPLICANT: Blanc, Veronique
 : APPLICANT: Blanche, Francis
 : APPLICANT: Crouzet, Joel
 : APPLICANT: Jacques, Nathalie
 : APPLICANT: Lacroix, Patricia
 : APPLICANT: Thibaut, Denis


```

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..474
; OTHER INFORMATION: /product= "Partie du gene Smbd"
US-08-510-646B-14

Query Match
Best Local Similarity 6.0%; Score 60.4; DB 3; Length 474;
Matches 184; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

Qy 199 TCCGACGCGCTTTGCCGCGCTTCTGCTGGAGAGCGAGGTTGCGTACATGTTA 258
Db 35 TGCCCTGGAGACACCCCGCGCGCGTGGCCACCCAGCGCGCGCGCGCTTCC 94

Qy 259 CCATCGAGCGGAGCTTACACAGTCCCTTTCGAGCGCTTTCGCAAGAACACACGACCT 318
Db 95 ACATCGAGCGCGCTTCCACGAGAAGCTGACCGCGCTTCCAGGCGCTGGACAGCGCC 154

Qy 319 CTTTCGTGTTCTTCTAGCTCGGTTCCGTTGCGGCTCATATTCGTCACAGCTGTTGAAG 378
Db 155 TGTTCATGTTGCTCCAGCGCGGTCGCGCGCTGCTCACCGCGCGCGCGCGCGCG 214

Qy 379 ACGTGTGCTATGTTGATGACCAATTCGAATTCGCAACCGACCTGAACCTGGAGGATATCATCG 438
Db 215 ACATCCCGTGGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274

Qy 439 GCTGCTTTGTCAATACGAGTGTATGCGAATCAACATAGATCATACGATACCTTTGGGA 498
Db 275 GCTTCTTCGTCACACCGCTGCTGCGCACCGACCTCGCGCGCGCGCGCGCGCGCG 334

Qy 499 CTTTGATCAACCAAGTCAAGCTACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 558
Db 335 AACTCGTGCGAGCGGTCGCGCGGTTGCGACCTTCGCGCGCGCTACACGACGAGGATGCGGT 394

Qy 559 TTGAGCGGTTGTTATCAGCACTACAGCGCTG 588
Db 395 TCGAAAGCTGTCGAGAGGTCACACCGG 424

```

```

RESULT 5
US-09-231-818-14
; Sequence 14, Application US/09231818
; Patent No. 6171846
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanc, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/231.818
; FILING DATE:
; PRIOR APPLICATION DATA:

```

```

;
; APPLICATION NUMBER: US/08/403.852
; FILING DATE: 10-MAY-1995
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03806.0054-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..474
; OTHER INFORMATION: /product= "Partie du gene Smbd"
US-09-231-818-14

```

```

Query Match
Best Local Similarity 6.0%; Score 60.4; DB 4; Length 474;
Matches 184; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

Qy 199 TCCGACGCGCTTTGCCGCGCTTCTGCTGGAGAGCGAGGTTGCGTACATGTTA 258
Db 35 TGCCCTGGAGACACCCCGCGCGCGTGGCCACCCAGCGCGCGCGCGCTTCC 94

Qy 259 CCATCGAGCGGAGCTTACACAGTCCCTTTCGAGCGCTTTCGCAAGAACACACGACCT 318
Db 95 ACATCGAGCGCGCTTCCACGAGAAGCTGACCGCGCTTCCAGGCGCTGGACAGCGCC 154

Qy 319 CTTTCGTGTTCTTCTAGCTCGGTTCCGTTGCGGCTCATATTCGTCACAGCTGTTGAAG 378
Db 155 TGTTCATGTTGCTCCAGCGCGGTCGCGCGCTGCTCACCGCGCGCGCGCGCGCG 214

Qy 379 ACGTGTGCTATGTTGATGACCAATTCGAATTCGCAACCGACCTGAACCTGGAGGATATCATCG 438
Db 215 ACATCCCGTGGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274

Qy 439 GCTGCTTTGTCAATACGAGTGTATGCGAATCAACATAGATCATACGATACCTTTGGGA 498
Db 275 GCTTCTTCGTCACACCGCTGCTGCGCACCGACCTCGCGCGCGCGCGCGCGCGCG 334

Qy 499 CTTTGATCAACCAAGTCAAGCTACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 558
Db 335 AACTCGTGCGAGCGGTCGCGCGGTTGCGACCTTCGCGCGCGCTACACGACGAGGATGCGGT 394

Qy 559 TTGAGCGGTTGTTATCAGCACTACAGCGCTG 588
Db 395 TCGAAAGCTGTCGAGAGGTCACACCGG 424

```

```

RESULT 6
US-09-103-840A-2/C
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.

```

```

RESULT 7
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

RESULTS

QY 352 TACACCAATTCGGAATCCGAACGACCTGAACCTGAGGATATCATCGCTCTTTGTCAA 451
 Db 113886 ATTTCCCATCGCCGCGCGCATCTGCGCTGGATAACCTTGTGGGCTTTTGTCAA 113945
 QY 452 TACGAGTGTATGGAATCAACATAGATCATCAGATACCTTTGGGACTTTGTATCAACCA 511
 Db 113946 CACCTTGTGTGGGTGCTACACCTGGCGGTGATCCCGAGCTTCGCCGAACTGCTGGGGCA 114005
 QY 512 AGTAAAGCTACGACGACGAGCATTCGAGACGAGGATATTCGGTTTGTAGCGCTGTGT 571
 Db 114006 GTGCGAGCGCGCAGCCTGGCGCTACGAAATCAAGACGTACTTTTCGAGGTGCTCGT 114065

RESULT 11
 US-08-222-617A-3
 : Sequence 3, Application US/08222617A
 : Patent No. 5882879
 : GENERAL INFORMATION:
 : APPLICANT: Veenstra, Annemarie E.
 : APPLICANT: Martin, Juan F.
 : APPLICANT: Garcia, Bruno D.
 : APPLICANT: Gutierrez, Santiago
 : APPLICANT: Barredo, Jose L.
 : APPLICANT: Von Doehren, Hans
 : APPLICANT: Palissa, Harriet
 : APPLICANT: Montenegro, Eduardo P.
 : TITLE OF INVENTION: A Method for Influencing Beta-Lactam
 : TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
 : TITLE OF INVENTION: Quantities of ACV Synthetase
 : NUMBER OF SEQUENCES: 27
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 : STREET: 300 South Wacker Drive
 : CITY: Chicago
 : STATE: Illinois
 : COUNTRY: USA
 : ZIP: 60606
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/222,617A
 : FILING DATE: 04-APR-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : REFERENCE/DOCKET NUMBER: 97,157
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 11601 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : ORIGINAL SOURCE:
 : ORGANISM: Acremonium chrysogenum
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 388..11526
 : OTHER INFORMATION: /function= "Enzyme"
 : OTHER INFORMATION: /product= "ACV Synthetase"
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 8050..8052
 : OTHER INFORMATION:
 : OTHER INFORMATION: /note= "NNN-GCC, AGU, AGC, UCU, UCC, UCA, or UCG;
 : OTHER INFORMATION: Xaa=Ala or Ser "
 : US-08-222-617A-3
 : Query Match 4.1%; Score 41.2; DB 2; Length 11601;
 : Best Local Similarity 60.9%; Pred. No. 0.0064;

Matches 67; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 QY 387 ATTGTCACCAATTCGGAATCCGAACGACCTGAACCTGAGGATATCATCGCTCTTT 446
 Db 7129 ATTGTCATTCGCTTGGCCACCGTAACCATCGGACTTTGAGTCGGTGTGTCGGCTCTTT 7188
 QY 447 GTCAATGCGAGTGTATCGGAATCAACATAGATCATCAGATACCTTTGG 496
 Db 7189 GTCAACTTGTCTCCCTCTCGGGGTCAACGTGTCTCAGTCGGACATTCATGG 7238
 RESULT 12
 US-08-222-617A-24
 : Sequence 24, Application US/08222617A
 : Patent No. 5882879
 : GENERAL INFORMATION:
 : APPLICANT: Veenstra, Annemarie E.
 : APPLICANT: Martin, Juan F.
 : APPLICANT: Garcia, Bruno D.
 : APPLICANT: Gutierrez, Santiago
 : APPLICANT: Barredo, Jose L.
 : APPLICANT: Von Doehren, Hans
 : APPLICANT: Palissa, Harriet
 : APPLICANT: Van Liempt, Henk
 : APPLICANT: Montenegro, Eduardo P.
 : TITLE OF INVENTION: A Method for Influencing Beta-Lactam
 : TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
 : TITLE OF INVENTION: Quantities of ACV Synthetase
 : NUMBER OF SEQUENCES: 27
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 : STREET: 300 South Wacker Drive
 : CITY: Chicago
 : STATE: Illinois
 : COUNTRY: USA
 : ZIP: 60606
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/222,617A
 : FILING DATE: 04-APR-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : REFERENCE/DOCKET NUMBER: 97,157
 : INFORMATION FOR SEQ ID NO: 24:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 11601 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : ORIGINAL SOURCE:
 : ORGANISM: Acremonium chrysogenum
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 388..11526
 : OTHER INFORMATION: /function= "Enzyme"
 : OTHER INFORMATION: /product= "ACV Synthetase"
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 8050..8052
 : OTHER INFORMATION:
 : OTHER INFORMATION: /note= "NNN-AGU, AGC, UCU, UCC, UCA, or UCG"
 : US-08-222-617A-24
 : Query Match 4.1%; Score 41.2; DB 2; Length 11601;
 : Best Local Similarity 60.9%; Pred. No. 0.0064;
 : Matches 67; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 QY 387 ATTGTCACCAATTCGGAATCCGAACGACCTGAACCTGAGGATATCATCGCTCTTT 446

Db
7129 ATTGGTATTCGGTTCCGCACCGTAACCATCCGGAATTGAGTCGGTGTTCGCCCTCCTTT 7188

QY
447 GTCAATACGACGAGTGATCGAATCAACATAGATCATCACGATACCTTTGG 496

Db
7189 GTCAACTTCCTCCCTCTCGGGGTACACGTTCTCAGTCGACATTCATGG 7238

RESULT 13

```

US-08-743-200-5
; Sequence 5, Application US/08743200
; Patent No. 5861260
; GENERAL INFORMATION:
;
; APPLICANT: THLIVERKIS, ANDREW
;
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
;
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
;
; NUMBER OF SEQUENCES: 102
;
; CORRESPONDENCE ADDRESS:
;

```

Query Match 3.3%; Score 33; DB 2; Length 330;
Best Local Similarity 69.2%; Pred. No. 0.38;
Matches 45; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

RESULT 14

US-08-452-655B-1/c
; Sequence 1, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVIERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATION
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,655B
 FILING DATE: 25-MAY-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/289,548
 FILING DATE: 12-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.49964
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8532 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: DP2.5(APC)
 PS-08-452-655B-1

| | | | | | | |
|----|-----------------------|--|----------------|-------|--------------|------------|
| | Query Match | 3.3%; | Score 33; | DB 1; | Length 8532; | |
| | Best Local Similarity | 55.8%; | Pred. No. 2.8; | | | |
| | Matches | 63; | Conservative | 0; | Mismatches | 50; Indels |
| | | | | | | 0; Gaps |
| | | | | | | 0; |
| QY | 880 | TTCCTTTGACTGATGCATTGTGACCTTGGAAAAATTGGATGTCTCAACGTCAAACATG | 939 | | | |
| | | | | | | |
| Dd | 8112 | TTCCTTTGCTGATTATCTTTTGAATCTTTAAGTTTGATTTGCCCTTTTTCTCAACACT | 8053 | | | |
| | | | | | | |
| QY | 940 | TGCACTATCCCGAGAAATCGAGCTTGGCTGATGCTTCCAGACCCCAGTCTCT | 992 | | | |
| | | | | | | |
| Dd | 8052 | GTCAAATCACCGGGGAGTATTACCTCTGGGAGATCTTCCAGATCTAGGATTGT | 8000 | | | |

RESULT 15

RESOL 13
US-08-450-582-1/c
; Sequence 1, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/450,582
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/452,655
 FILING DATE: 25-MAY-1995
 APPLICATION NUMBER: US 08/289,548
 FILING DATE: 12-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.49964
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8532 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: DP2.5(APC)
 US-08-450-582-1

Query Match 3.3%; Score 33; DB 3; Length 8532;
 Best Local Similarity 55.8%; Pred. No. 2.8;
 Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 880 TTCCTTTGACTGATGCATTGTGACTCTTGAAAAATTGGATGTTCTCAACGTCACACATG 939
 DB 8112 TTGTTTGGCTGATTATCTTTTGAATCTTTAATGTTGGATTGGCTTTTCTGAAACACT 8053
 QY 940 TCGACTATCCCGAGAATCGAGCTTGGCTGTGCTTCCAGACCCCAAGTCTCT 992
 DB 8052 GTCAATCACGGGGGAGGATATTACCTGTGGAGATCTTCAGATCTAGATTGT 8000

Search completed: June 4, 2003, 21:50:58
 Job time : 93 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 20:51:29 ; Search time 170 Seconds
(without alignments)
8240.805 Million cell updates/sec

Title: us-09-482-788-1_COPY_4000_5000

Perfect score: 1001

Sequence: 1 caattgatgtcttgacgc.....ccccagctctgtacc 1001

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2.6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2.6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2.6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2.6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2.6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2.6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2.6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2.6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2.6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2.6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2.6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2.6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2.6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----|---------------------|
| 1 | 75.6 | 7.6 | 13029 | 10 | US-09-815-242-4052 |
| 2 | 71.8 | 7.2 | 3471 | 10 | US-09-974-300-2167 |
| 3 | 65.8 | 6.6 | 88421 | 9 | US-09-976-059-1 |
| 4 | 60.4 | 6.0 | 7347 | 10 | US-09-815-242-7773 |
| 5 | 56 | 5.6 | 7158 | 10 | US-09-974-300-2171 |
| 6 | 47.2 | 4.7 | 6455 | 10 | US-09-974-300-2166 |
| 7 | 47.2 | 4.7 | 8268 | 10 | US-09-974-300-2169 |
| 8 | 41.6 | 4.2 | 693 | 9 | US-10-123-155-406 |
| 9 | 34.8 | 3.5 | 558 | 9 | US-09-918-995-31196 |
| 10 | 34.8 | 3.5 | 558 | 9 | US-09-918-995-28625 |
| 11 | 34.2 | 3.4 | 269 | 10 | US-09-923-876-5113 |
| 12 | 33.8 | 3.4 | 175561 | 9 | US-10-017-721-3 |
| 13 | 33.3 | 3.3 | 10383 | 10 | US-09-960-253-181 |
| 14 | 32.8 | 3.3 | 6827 | 10 | US-09-982-610-17 |
| 15 | 32.6 | 3.3 | 1290 | 10 | US-09-815-242-7834 |
| 16 | 32.2 | 3.2 | 2277 | 9 | US-10-189-971-21 |
| 17 | 32.2 | 3.2 | 3173 | 9 | US-10-189-971-5 |
| 18 | 32.2 | 3.2 | 3753 | 9 | US-10-189-971-15 |
| 19 | 32.2 | 3.2 | 4026 | 9 | US-10-189-971-23 |

| | | | | | |
|------|------|-----|-------|----|---------------------|
| c 20 | 32.2 | 3.2 | 4536 | 9 | US-10-189-971-9 |
| c 21 | 32.2 | 3.2 | 4710 | 9 | US-10-189-971-11 |
| c 22 | 32.2 | 3.2 | 4884 | 9 | US-10-189-971-1 |
| c 23 | 32.2 | 3.2 | 5776 | 9 | US-10-189-971-25 |
| c 24 | 31.8 | 3.2 | 266 | 10 | US-09-923-876-5115 |
| c 25 | 31.8 | 3.2 | 285 | 10 | US-09-815-242-9092 |
| c 26 | 31.6 | 3.2 | 913 | 9 | US-10-091-504-2158 |
| c 27 | 31.6 | 3.2 | 913 | 10 | US-09-764-869-2158 |
| c 28 | 31.6 | 3.2 | 1160 | 9 | US-10-123-155-234 |
| c 29 | 31.6 | 3.2 | 1590 | 9 | US-09-298-5238-75 |
| c 30 | 31.4 | 3.1 | 450 | 7 | US-08-781-986A-1027 |
| c 31 | 31.4 | 3.1 | 3122 | 10 | US-09-908-500A-1 |
| c 32 | 31.4 | 3.1 | 3603 | 9 | US-09-742-153-1 |
| c 33 | 31.2 | 3.1 | 478 | 9 | US-10-123-155-138 |
| c 34 | 31.2 | 3.1 | 4111 | 9 | US-09-375-248-1 |
| c 35 | 31.2 | 3.1 | 4416 | 9 | US-10-201-386-1 |
| c 36 | 31.2 | 3.1 | 4425 | 10 | US-09-982-610-31 |
| c 37 | 31.2 | 3.1 | 9108 | 10 | US-09-982-610-45 |
| c 38 | 31 | 3.1 | 49744 | 10 | US-09-927-091-4 |
| c 39 | 30.6 | 3.1 | 802 | 9 | US-10-184-644-312 |
| c 40 | 30.6 | 3.1 | 802 | 9 | US-10-184-644-312 |
| c 41 | 30.6 | 3.1 | 1254 | 9 | US-09-822-846-75 |
| c 42 | 30.4 | 3.0 | 444 | 9 | US-10-123-155-498 |
| c 43 | 30.4 | 3.0 | 1314 | 10 | US-09-974-300-1682 |
| c 44 | 30.4 | 3.0 | 1711 | 10 | US-09-867-550-1379 |
| c 45 | 30.4 | 3.0 | 1799 | 10 | US-09-824-567-1 |

ALIGNMENTS

RESULT 1

- US-09-815-242-4052
- Sequence 4052, Application US/09815242
- Patent No. US20020061569A1
- GENERAL INFORMATION:
- APPLICANT: Haselbeck, Robert
- APPLICANT: Ohlsen, Kari L.
- APPLICANT: Zyskind, Judith W.
- APPLICANT: Wall, Daniel
- APPLICANT: Trawick, John D.
- APPLICANT: Carr, Grant J.
- APPLICANT: Yamamoto, Robert T.
- APPLICANT: Xu, H. Howard
- TITLE OF INVENTION: Identification of Essential Genes in
- FILE REFERENCE: ELITRA 011A
- CURRENT APPLICATION NUMBER: US/09/815,242
- CURRENT FILING DATE: 2001-03-21
- PRIOR APPLICATION NUMBER: 60/191,078
- PRIOR FILING DATE: 2000-03-21
- PRIOR APPLICATION NUMBER: 60/206,848
- PRIOR FILING DATE: 2000-05-23
- PRIOR APPLICATION NUMBER: 60/207,727
- PRIOR FILING DATE: 2000-05-26
- PRIOR APPLICATION NUMBER: 60/242,578
- PRIOR FILING DATE: 2000-10-23
- PRIOR APPLICATION NUMBER: 60/253,625
- PRIOR FILING DATE: 2000-11-27
- PRIOR APPLICATION NUMBER: 60/257,931
- PRIOR FILING DATE: 2000-12-22
- PRIOR APPLICATION NUMBER: 60/269,308
- PRIOR FILING DATE: 2001-02-16
- NUMBER OF SEQ ID NOS: 14110
- SOFTWARE: FastSeq for Windows Version 4.0
- SEQ ID NO 4052
- LENGTH: 13029
- TYPE: DNA
- ORGANISM: Pseudomonas aeruginosa

US-09-815-242-4052

Query Match 7.6%; Score 75.6; DB 10; Length 13029;
Best Local Similarity 45.6%; Pred. No. 4.5e-14;


```
Query Match          6.6%; Score 65.8; DB 9; Length 88421;
Best Local Similarity 48.5%; Pred. No. 3.4e-10;
Matches 181; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
```

RESULT 4

```

; LIFE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7347)
US-09-815-242-7773

```

Query Match 6.0%; Score 60.4; DB 10; Length 7347;

[illegible]

```

RESULT 5
US-09-974-300-2171
; Sequence 2171, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2171
; LENGTH: 7158
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2171

```

| Query Match | 5.6% | Score 56 | DB 10 | Length 7158 |
|-----------------------|----------------|---|----------|-------------|
| Best Local Similarity | 47.7% | Pred. No. 1.4e-07 | | |
| Matches 164 | Conservative 0 | Mismatches 180 | Indels 0 | Gaps 0 |
| QY | 243 | GGTGGTACATCTTACCATCGAGCGAGCTCTACCAGTCCCTTCGAGCCTTCTGCAAC | 302 | |
| Db | 3820 | GGTACATTGATCTCAGCATTGAGCAACACTGCACGGCGCCTGTTTCGATTTGTCCTCCGC | 3879 | |
| QY | 303 | GAACACACACGACCTCTTTTCGTGTTCTTCTAGCTGCGGTTCGGTCCGCTCATATATCGT | 352 | |
| Db | 3880 | AGACCCGGAGTCAGCATGTTTATGATTTTGCACTCGGCTCTTCCCGCACTACTGCACACGG | 3939 | |
| QY | 363 | CTCACAGCTGTTTGAAGACCGCTGTCTATGGTGTACACCAATTCGGAATCGCAACGACCTGAA | 422 | |
| Db | 3940 | CTCGGAGCAGGCCATCATATTCGCTCGCAGCCCGATAGCGGGAAGAACGATGATGCT | 3999 | |
| QY | 423 | CTCGAGGATATCATCGGCTGCTTTTGTCATATACCAGTGTATGCGAATCAACATAGATCAT | 482 | |
| Db | 4000 | TTAGCGGAATTCGCGATTGTTTGTATACCGTTTGTGCTTCGACACAGATACATCGGGT | 4059 | |
| QY | 483 | CACGATACCTTTGGGACTTTTGATCAACCAAGTCAAGGCTACGACGACGACGATTCGAG | 542 | |

```

Db      4060  AATCCGAGTTTCGGGAGCTTCTCAACAGAGAGTCAGAAAAGCTTAACCTTCGACGGTATGAG 4111
QY      543   AACGAGGATATTCGGTTTGAGCGCGTTGTATCATCAGCACTACAGCC 586
Db      4120  CATCAGGACCTTCGGTTTCGAGCGCTCTGTTGAAGTGCCTTAATCC 4163

RESULT 6
US-09-974-300-2166
; Sequence 2166, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berkay, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2166
; LENGTH: 6465
; TYPE: DNA
; ORGANISM: Bacillus licheniformis.
US-09-974-300-2166

```

| Query Match | 4.7% | Score 47.2; | DB 10; | Length 6465; |
|-----------------------|-----------------|---|-----------|--------------|
| Best Local Similarity | 45.6%; | Pred. No. 0.00013; | | |
| Matches 166; | Conservative 0; | Mismatches 198; | Indels 0; | Gaps 0; |
| Qy | 199 | TCGCGACGAGCTTTGCCGCGCTTGCACCTCTGCTCGAGACGCGAGCTTGGTACATGTTA | 258 | |
| Db | 1534 | | | |
| Qy | 259 | CCATCGACGCGAGCTCTACCACTCCCTTGGAGCCCTTCTGCAACGAACACACACGACCT | 318 | |
| Db | 1594 | GAATCAATCAAAACATGACGGCAGAGCTTCAAAAGAGCTGCTTCTGAAGCCGATGCGACAC | 1653 | |
| Qy | 319 | CTTTCGTCGTTCTCTAGCTGCGCTTCCGTGCCCTCATTTATCGTCTCACAGCTGTTGAAG | 378 | |
| Db | 1654 | TTTATATGTTCCCTGTTGGCGCGCTTTAAATTTTGTCTGTCCAAAGTATGCGTCTCAGGAAG | 1713 | |
| Qy | 379 | ACGCTGTCATTGGTATACACCAATTTGCCAATTCGCAACCGAGCTGAACCTGGAGGATATCATCG | 438 | |
| Db | 1714 | ATTTAATCGTCGGATCGCCTGTGCGCGGAGAACCCATCCGGATCTGCACAACTTCCGG | 1773 | |
| Qy | 439 | GCCTGCTTTGCAATACGCAGTGTATCGGAATCAACATAGATCATTCACGATACCTTTGGGA | 498 | |
| Db | 1774 | GCATGTTTGTCAATACGCTGGCGCTCCGGAATGCTCCCGAAGGGGAAAAAGCTTTAAGG | 1833 | |
| Qy | 499 | CTTTTGTCAACCAAGTCAAGGCTACGACGACAGCAGCATTTTCGAGAACGAGGATATTCGCT | 558 | |
| Db | 1834 | AATTTTTGCAGGAAGTCAAGAGACAAGTCTTCAGGGCTTTGCCAAGCAGGACTATCCGC | 1893 | |
| Qy | 559 | TTGA 562 | | |
| Db | 1894 | TTGA 1897 | | |

```

RESULT 7
US-09-974-300-2169
; Sequence 2169, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

```


140 GCAGGAGAGCAACTCACTACTGGAGAGCACTCAAGACTTCTCCCGCAAGAT 199
QY
287 GCAGAAAACAAAAAATCAACAAAGAGCCCTCTGCTCTCCCGGGAAGAC 346
Db
200 CCCGACCGACTTTGGCCGCCCTGTCTCTGGAGAGC 241
QY
347 CCTGACCGGTGATACATAGCCCTGGTGTCTCTGCCAGCCAGCC 388
Db

```

RESULT 10
US-09-918-995-28625
; Sequence 28625, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 204111-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28625
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(558)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-28625

```

```

RESULT 11
; US-923-876-5113
; Sequence 5113, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 5113
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700456379H1
; NAME/KEY: unsure

```

```

; LOCATION: 149, 190
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-5113

Query Match          3.4%; Score 34.2; DB 10; Length 269;
Best Local Similarity 54.5%; Pred. No. 0.45;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 629 TTTTGCAGTGCACCTACAGAAGGACCTTTGGAAGATTCAAGTTCAGGGTCTCGAGTCCGT 688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102 TTTTCACGGTGGAGTTCCCCGACTCGCTGGCACCGACGAGTSCAAGGNTCTCGAGTCGGT 161

QY 689 ACTGTGCCTACGAAGGTCACACTCGATTTGCACATGGAGTTCGATCTGTTTCAGAAAC 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 162 ACTTCCGCCAAAGCCTTCGTCCAAGCTNCCGCACATGGAGATCGATGAATCCGAGGGGC 221

QY 749 C 749
    |
Db 222 C 222

```

```

RESULT 12
US-10-017-721-3
; Sequence 3, Application US/10017721
; Publication No. US20030096248A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Daley, George
; APPLICANT: Bolk, Stacey
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: MMI-003
; CURRENT APPLICATION NUMBER: US/10/017.721
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/317,033
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: US 60/330,248
; PRIOR FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 175561
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-017-721-3

```

```

RESULT 13
US-09-960-253-181/c.
; Sequence 181, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 181
; LENGTH: 10383
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9089, 9347, 9453, 9519, 10205
; OTHER INFORMATION: n = A,T,C or G
US-09-960-253-181

Query Match          3.3%; Score 33; DB 10; Length 10383;
Best Local Similarity 55.8%; Pred. No. 12;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 880 TTCCCTTGACTGATGCTGACTCTTGAAGAAATGGATGTTCTCAACGTCACAACTG 939
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8150 TTGTTTGCTGATATCTTTGAATCTTTATGTTGGATTTGCTTTTCTGAACACT 8091

QY 940 TCGACTATCCCGAGAAATCGAGCTGGCTGGTGTCTTCAGACCCCAAGTCTCT 992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8090 GTCAATCACCGGGGAGTATTACCTGTGGGAGATCTTCAGATCTAGGATTGT 8038

RESULT 14
US-09-982-610-17
; Sequence 17, Application US/09982610
; Patent No. US20020146420A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; BENNETT, BRIAN D.
; GOEDDEL, DAVID
; MATTHEWS, WILLIAM
; TSAI, SIAO PING
; WOOD, WILLIAM I.
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/982,610
; FILING DATE: 17-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/446,648
; FILING DATE: 1996-MAY-23
; APPLICATION NUMBER: 08/222616
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P0821P3PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6827 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```

```

US-09-982-610-17

Query Match          3.3%; Score 32.8; DB 10; Length 6827;
Best Local Similarity 52.1%; Pred. No. 11;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 348 GCCGCTCATTTATGCTCTCACAGCTGTTGAAGAGCGCTGCTTGTCTCAATAGGAGTGTATGCGAAT 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2486 GTCACCTCTCGTGGTGTACACCTTGTCTGAAGATGCTTTCAGGGGCCATCCACTTTCAGGGG 2545

QY 408 CGCAACCGGACCTGAACTGGAGGATATCATCGGCTGCTTGTCTCAATAGGAGTGTATGCGAAT 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2546 CAGCCGGGCACTGCCCTTCGGGACGTAGTCGGGGTCTTGTGTAGATGTCCCGGCAAGGCC 2605

QY 468 ATCAACATAGATCATCACGA 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2606 AAGTCACAGATCTTCACCA 2625

RESULT 15
US-09-815-242-7834/c
; Sequence 7834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7834
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-09-815-242-7834

Query Match          3.3%; Score 32.6; DB 10; Length 1290;
Best Local Similarity 48.2%; Pred. No. 4.4;
Matches 92; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 199 TCCCGACCGGACTTTGCCCGCCCTGACACTTCTGTCTGGAGACGCGAGTTGCGTACATGTTA 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 782 TCCAGGTCTGCTACTTGGCGCTCTTGAAGAACTCGCTGGAGCGGAGTCCAGGCGCAGGGTC 723

QY 259 CCATCGACGCGGAGCTCTACCACTCCCTTCGACCTTCTCAACGACGACGACCT 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 722 ACGTCTCGGCCCGGCTTGTAGCGGCGCTTCTTCAGCGGCTTCGGCGATGCGCGCAGGGCG 663

```

319 CTTCGTCGTTCTCTAGCTCCGTTCCGTCGCCCTCATTTATCGTCTFCACAGCTGTTGAAG 378
 662 TCTCTGTTGGACGACAGGTTCCGGCGGAAGCCCGCTTCGTCACCGACGCGGTGTTTCAGG 603
 379 ACGCTGTCAATT 389
 602 CCACGGGCGCTT 592
 QY
 Db

Search completed: June 4, 2003, 23:54:08
Job time : 175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:53:27 ; Search time 1708.5 Seconds
(without alignments)
9488.839 Million cell updates/sec

Title: US-09-482-788-1_COPY_4000_5000
Perfect score: 100
Sequence: 1 caattgatgttgcagcgc.....cccaagtctgtctaacccc 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```
EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 66.4 | 6.6 | 438 | 17 | AQ989462 Rfc000006 |
| 2 | 41.8 | 4.2 | 637 | 17 | AQ990256 Rfc000996 |
| 3 | 40.4 | 4.0 | 797 | 17 | AQ991640 Rfc01385F |
| 4 | 38.6 | 3.9 | 1145 | 12 | BF036321 60145954 |
| 5 | 38 | 3.8 | 786 | 10 | BE130762 L48-973T3 |
| 6 | 37.8 | 3.8 | 791 | 13 | BM012777 603637492 |

| | | | | | | | |
|---|----|------|-----|------|----|----------|---------------------|
| c | 7 | 37.2 | 3.7 | 492 | 12 | BF653686 | BF653686 277599 MA |
| c | 8 | 37.2 | 3.7 | 493 | 12 | BF653730 | BF653730 277659 MA |
| c | 9 | 36.8 | 3.7 | 280 | 12 | BF643966 | BF643966 NF010A02E |
| | 10 | 36.8 | 3.7 | 483 | 12 | BF646516 | BF646516 NF076F01E |
| | 11 | 36.8 | 3.7 | 586 | 12 | EG453877 | EG453877 NF095E06L |
| | 12 | 36.8 | 3.7 | 662 | 12 | BF644804 | BF644804 NF042H10E |
| | 13 | 36.8 | 3.7 | 668 | 10 | AW689867 | AW689867 NF025C03S |
| c | 14 | 36.6 | 3.7 | 591 | 9 | AI737351 | AI737351 606039D04 |
| | 15 | 36.6 | 3.7 | 638 | 10 | AW056085 | AW056085 660004A04 |
| | 16 | 36.4 | 3.6 | 939 | 17 | CNS00CNG | AL059400 Drosophil1 |
| c | 17 | 36.2 | 3.6 | 816 | 10 | BE641516 | BE641516 Cri12_3_H2 |
| c | 18 | 36.2 | 3.6 | 840 | 10 | BE640820 | BE640820 Cri12_1_H0 |
| | 19 | 35.6 | 3.6 | 383 | 17 | BH475809 | BH475809 BOHDE17TR |
| | 20 | 35.6 | 3.6 | 736 | 17 | BH179075 | BH179075 O13_D_05- |
| | 21 | 35.6 | 3.6 | 736 | 17 | CNS07L8T | AL616031 T3 end of |
| | 22 | 35.6 | 3.6 | 761 | 17 | BH557436 | BH557436 BOGNK42TF |
| | 23 | 35.4 | 3.5 | 208 | 14 | BM791536 | BM791536 K-EST0071 |
| c | 24 | 35.4 | 3.5 | 384 | 14 | T65025 | T65025 YC75C05.s1 |
| c | 25 | 35.4 | 3.5 | 600 | 14 | BQ783688 | BQ783688 fab32f08. |
| c | 26 | 35.4 | 3.5 | 1101 | 17 | CNS00LT2 | AL078714 Drosophil1 |
| c | 27 | 35.2 | 3.5 | 729 | 17 | BH679719 | BH679719 BOHVX54TF |
| c | 28 | 35.2 | 3.5 | 748 | 17 | AG090994 | AG090994 Pan trogl |
| c | 29 | 35.2 | 3.5 | 991 | 12 | BG821150 | BG821150 602724546 |
| | 30 | 35 | 3.5 | 496 | 17 | A2323994 | AZ323994 1M0045A18 |
| | 31 | 35 | 3.5 | 565 | 10 | AW562597 | AW562597 660066H04 |
| c | 32 | 35 | 3.5 | 844 | 10 | BE640899 | BE640899 Cri12_1_L0 |
| c | 33 | 35 | 3.5 | 938 | 14 | BQ619362 | BQ619362 RNOSEQH0 |
| | 34 | 35 | 3.5 | 1031 | 11 | AY104862 | AY104862 Zea mays |
| | 35 | 34.8 | 3.5 | 248 | 12 | BG050913 | BG050913 FM1_72.G0 |
| c | 36 | 34.8 | 3.5 | 295 | 9 | AA627760 | AA627760 nq48C02.S |
| | 37 | 34.8 | 3.5 | 340 | 9 | AI586493 | AI586493 486051H08 |
| | 38 | 34.8 | 3.5 | 638 | 10 | AW958464 | AW958464 EST370534 |
| | 39 | 34.8 | 3.5 | 646 | 10 | AW953398 | AW953398 EST365468 |
| | 40 | 34.8 | 3.5 | 993 | 14 | BQ064890 | BQ064890 AGENCOURT |
| c | 41 | 34.8 | 3.5 | 1101 | 17 | CNS00PXE | AL071370 Drosophil1 |
| c | 42 | 34.6 | 3.5 | 674 | 9 | AI491368 | AI491368 486037H10 |
| | 43 | 34.6 | 3.5 | 675 | 9 | AL692518 | AL692518 AL692518 |
| c | 44 | 34.6 | 3.5 | 693 | 9 | AL450949 | AL450949 AL450949 |
| c | 45 | 34.6 | 3.5 | 709 | 12 | BG593086 | BG593086 EST491764 |

ALIGNMENTS

RESULT 1
AQ989462 438 bp DNA linear GSS 14-AUG-2000
LOCUS Rfc000006 Photorhabdus luminescens strain W14 M13 library
DEFINITION Photorhabdus luminescens genomic clone PLG000006, DNA sequence.
ACCESSION AQ989462
VERSION AQ989462.1 GI:9648056
KEYWORDS GSS.
SOURCE Photorhabdus luminescens.
ORGANISM Photorhabdus luminescens.
REFERENCE 1 (bases 1 to 438)
AUTHORS ffrench-Constant, R.H., Waterfield, N., Buxland, V., Perna, N.T., Daborn, P.J., Bowen, D. and Blattner, F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens W14: potential implications for virulence
JOURNAL Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE 20378633
COMMENT Contact: ffrench-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bssrfc@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic

Acids Res.

Seq primer: M13 Forward
Class: shotgun.

FEATURES
Location/Qualifiers
1..438

source /organism="Photorhabdus luminescens"

/strain="W14"

/db_xref="taxon:29488"

/clone="PLG00006"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/dev_stage="primary phase variant"

/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

BASE COUNT 92 a 93 c 149 g 104 t

ORIGIN

Query Match 6.6%; Score 56.4; DB 17; Length 438;

Best Local Similarity 55.1%; Pred. No. 6.8e-09;

Matches 130; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 351 GCTCATATTCGCTCACAGCTGTTGAAGACGCTGTCAATGGTACACCAATTCGAATCGC 410

Db 7 GTTCTGCGCTGTCGGTCAAGAGATATGTCGCGACACCGAGTCGGGTCGC 66

QY 411 AACCGACTGAATCGAGGATATCATCGGCTGTTGTCAATACGCGAGTGTATCGCAATC 470

Db 67 AATCGTCAGGAAGTAGAACCAATGATGGCTTCTTTGTCATACAGCTGGCGTCGCGATA 126

QY 471 AACATAGATCATACGATACCTTTGGACCTTTGATCAACCAAGTCAAGGTACGACGACA 530

Db 127 GATGATTCGATGAGCTGATGTAACCGAGTGTTCGGCGCAATCCGGCAACCGCATG 186

QY 531 GCAGCATTCGAGAACGAGGATATTCGCTTTGAGCGCGTGTATCAGCACTACAGCC 586

Db 187 GCGCGCAGGATCAGGATTTACCGTTTGAACAGGTGTGGAGATCGTACACCC 242

RESULT 2

LOCUS

AO990256

DEFINITION

Photorhabdus luminescens strain W14 M13 library

Photorhabdus luminescens genomic clone PLG00996, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

1 (bases 1 to 637)

ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,

Daborn, P.J., Bowen, D., and Blattner, F.R.

A genomic sample sequence of the entomopathogenic bacterium

Photorhabdus luminescens W14: potential implications for virulence

Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)

20378633

Contact: ffrench-Constant RH

Department of Biology and Biochemistry

University of Bath

South Building, Bath BA2 7AY, UK

Tel: (44) 1225 826521

Fax: (44) 1225 826779

Email: bssrfce@bath.ac.uk

This is one of 2,122 random reads from the M13 library. For

annotation of identified clones (BLASTX, BLASTN and mapping to E.

coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic

Acids Res.

Seq primer: M13 Forward

Class: shotgun.

FEATURES

source

1..637

/organism="Photorhabdus luminescens"

Location/Qualifiers

1..797

/organism="Photorhabdus luminescens"

/strain="W14"

/db_xref="taxon:29488"

/clone="PLG01385F"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/dev_stage="primary phase variant"

/strain="W14"

/db_xref="taxon:29488"

/clone="PLG00996"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/dev_stage="primary phase variant"

/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

BASE COUNT 166 a 124 c 134 g 211 t 2 others

ORIGIN

Query Match 4.2%; Score 41.8; DB 17; Length 637;

Best Local Similarity 48.9%; Pred. No. 0.26;

Matches 112; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 358 ATCGTGTTCACAGCTGTTGAAGACGCTGTCAATGGTACACCAATTCGAATCGCAACCGAC 417

Db 10 ATCGTTATTCACAGGATGAAGACATCGTTATTTGTTAGTCCATTTGCCAATCGTCACAATA 69

QY 418 CTGAACCTGGAGGATATCATCGGCTGCTTTGTCAATACGCACTGTATCGCAATCAACATAG 477

Db 70 TTGATGAACATGTCGCTGTTGGTTTATTTATCAATTTATTAACGCTTCGGTTTTCGCTTTA 129

QY 478 ATCATCAGGATACCTTTGGGACTTTGATCAACCAAGTCAAGGCTACGACGACGACGAT 537

Db 130 ACAGGAGGATTAATTTCAAGTCAATTTGTTTCAAGCGCGCATATATTTCTTGATGCCT 189

QY 538 TCAGAACGAGGAGGATATTCGCTTTGAGCGCGTGTGATCAGCACTACAGCC 586

Db 190 ATCAAAATGCTGATTTGCCAATTTGATGAGATTTGTCGATGTGGTACAGCC 238

RESULT 3

AO991640

LOCUS

DEFINITION

Photorhabdus luminescens strain W14 M13 library

Photorhabdus luminescens genomic clone PLG01385F, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

1 (bases 1 to 797)

ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,

Daborn, P.J., Bowen, D., and Blattner, F.R.

A genomic sample sequence of the entomopathogenic bacterium

Photorhabdus luminescens W14: potential implications for virulence

Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)

20378633

Contact: ffrench-Constant RH

Department of Biology and Biochemistry

University of Bath

South Building, Bath BA2 7AY, UK

Tel: (44) 1225 826521

Fax: (44) 1225 826779

Email: bssrfce@bath.ac.uk

This is one of a selected subset of flipped clones from the M13

library. For annotation of identified clones (BLASTX, BLASTN and

mapping to E. coli K12 genome) please see ffrench-Constant et al.

2000, Nucleic Acids Res.

Seq primer: M13 Reverse

Class: shotgun.

FEATURES

source

1..797

/organism="Photorhabdus luminescens"

/strain="W14"

/db_xref="taxon:29488"

/clone="PLG01385F"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/dev_stage="primary phase variant"

1

.....

```

Db      714  GGCTTTTAATGAAGCTATTAAATGAAGGAGCTATAGCCAAATGGAGAAATTTTCGA 773
Qy      794  AA 795
        ||
Db      774  AA 775

RESULT 6
BM012777/c
LOCUS   603637492F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5459759 5',
        mRNA sequence.
ACCESSION BM012777
VERSION   BM012777.1 GI:16527131
KEYWORDS EST.
SOURCE    human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 791)
          NIH-MGC http://mgc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapsb-re@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCMI963 row: g column: 24
          High quality sequence stop: 285.
FEATURES             Location/Qualifiers
     source           1..791
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="IMAGE:5459759"
                     /clone_lib="NIH_MGC_47"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOT87; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCAAGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT          215 a 135 c 250 g 191 t
ORIGIN
Query Match          3.8%; Score 37.8; DB 13; Length 791;
Best Local Similarity 57.0%; Pred. No. 5;
Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy      505  TCAACCAAGTCAAGGCTACGACGACGACGATTCGAGACGAGGATATTCCTGTTGAGC 564
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      568  TCAACACATGTCAAGGCTCCGTCACCGTTAGCACTCTGACGAGACATACCTGAGTAGC 509
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      565  CGCTTGTATCAGACATACACCTGGATCCAGAGATCTGTGCAAGCAGACACCTCTCGCACAC 624
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      508  ACTCTTTCTGCTGGTACCTTCCACTTTTCACTCCCTTTTCAATCACCTCTCTGGAAGAT 449
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      625  T 625
        |
Db      448  T 448

RESULT 7
BF653686/c
LOCUS   277599 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF653686
VERSION   BF653686.1 GI:11918818
KEYWORDS EST.
SOURCE    cow.
ORGANISM Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
          Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 492)
          Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
          Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
          , G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
          Perlea, G., Holt, I.I., Karamycheva, S., Liang, F., Quackenbush, J. and
          Keele, J.W.
          Sequence evaluation of four pooled-tissue normalized bovine cDNA
          libraries and construction of a gene index for cattle
          Genome Res. 11 (4), 626-630 (2001)
          21180013
          Contact: Smith TPL
          USDA, ARS, US Meat Animal Research Center
          PO Box 166, Clay Center, NE 68933-0166, USA
          Tel: 402 762 4366
          Fax: 402 762 4390
          Email: smith@email.marc.usda.gov
          Single pass sequencing. Bases called and alt_trimmed with phred
          v0.980904.e. Vector identified by cross_match with the -minscore 18
          and -mismatch 12 options.
          PCR Primers
          FORWARD: AGGAACACGCTATGACCAT
          BACKWARD: GTTTTCCCAGTCACGACG
          Plate: 66 row: F column: 24
          Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source           1..492
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendinosus muscle, and fetal
                     longissimus muscle."
BASE COUNT          112 a 106 c 88 g 186 t
ORIGIN
Query Match          3.7%; Score 37.2; DB 12; Length 492;
Best Local Similarity 51.9%; Pred. No. 5.6;
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy      674  GGGTCTCGAGTCGCTACCTGTGCTAGCAAGCGTACACTCGATTTGACATGGAGTTCCA 733
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      425  GGCCCTGGAGTCAAAATGAGCAGACGACAAAGGCTACACAGTTTACCAGAGAATCCA 366
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      734  TCTGTTTCAAGAAACCGACAGCCCTTAAAGGTAGCGTCAACTTTGCCGATGAGCTGTCAA 793
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      365  CTGGTCATAGAAACACCCCTTTTCCAAACAAGACAACTCTACACATGACATCACA 306
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      794  AATGGAGACTGTTGAAATGTCGTCAGAGTATTTCTTTGAGAT 835
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      305  GATGGGCACTACTGAAATTTGGATTAATAATATTTCTTTGCAAT 264
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
BF653730/c
LOCUS   277659 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF653730
VERSION   BF653730.1 GI:11918862
KEYWORDS EST.
SOURCE    cow.

```

ORGANISM

Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 493)

REFERENCE
AUTHORS

Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keefe, J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine CDNA
libraries and construction of a gene index for cattle

JOURNAL
MEDLINE
COMMENT

Genome Res. 11 (4), 626-630 (2001)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt.trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCCGTCACGACG

Plate: 66 row: P column: 24

Seq primer: ATTAGTGACACTATAG.

FEATURES

Location/Qualifiers

source

1..493

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 3BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

BASE COUNT
ORIGIN

112 a 106 c 89 g 186 t

Query Match

Best Local Similarity 3.7%; Score 37.2; DB 12; Length 493;

Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY

674 GGCTCTCGAGTCGTCACCTGCTAGCAAGCGTACACTCGATTTCACATGGAGTTCCA 733

DB

426 GGCCCTGGAGTCAAAATGAGCAGAGCAAGGCTACAGAGTTTACCAGAGAAATGCA 367

QY

734 TCTGTTTCAAGAACCGACAGCGCTTAAAGTAGCTCAACTTTGCCGATGAGCTGTTCAA 793

DB

366 CTGCTCATAGAAACACCCCTTTTCCAACACAGACAACTCTACACATGACATCACCA 307

QY

794 AATGGAGACTGTGAAATGCTGTCAGAGTATTTCTTTGAGAT 835

DB

306 GATGGTCACTACTGAAATGGATTAATATATTCTTTGCAAT 265

RESULT 9

BF643966

LOCUS

NF010A02EC1F1006 Elicited cell culture Medicago truncatula CDNA
clone NF010A02EC 5', mRNA sequence.

ACCESSION

BF643966

VERSION

BF643966.1

KEYWORDS

EST.

SOURCE

barrel medic.

ORGANISM

Medicago

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 483)

Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,

Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation -

Center for Medicago Genomics Research

Unpublished (2000)

Contact: Dixon RA

Plant Biology Division

The Samuel Roberts Noble Foundation

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 280)

Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,

Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation -

Center for Medicago Genomics Research

Unpublished (2000)

Contact: Dixon RA

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7302

Fax: 580 221 7380

Email: radixon@noble.org

Insert Length: 280 Std Error: 0.00

Plate: 010 row: A column: 02

Seq primer: TCACACAGGAACACAGCTATGAC.

Location/Qualifiers

1..280

/organism="Medicago truncatula"

/db_xref="taxon:3880"

/clone_lib="NF010A02EC"

/tissue_type="Cell cultures derived from root tissues"

/dev_stage="Cell suspensions were subcultured every 14

days. Cells were induced six days after subculture"

/note="Vector: Lambda Zap; Cells were induced with yeast

cell wall extracts equivalent to 50ug/ml glucose in the

final concentration. Samples were taken at 0.5, 1, 12 and

24 hours after induction. Equal amounts of RNA from each

time point were pooled and used for mRNA isolation."

74 a 90 c 45 g 71 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

71; Conservative

0; Mismatches

57; Indels

0; Gaps

0;

QY

283 CCCTTCGAGCTTCTGCACGACACACACGACCTCTTTCGTGTTTCTTAGCTCGGT 342

DB

96 GCGGACCAACCTCCGCAACCAACGACAAATCGCTTCGTGTTTCTTAGCTCGGT 155

QY

343 TCGTGGCGCTCATTCGTCTCACAGCTGTTGAGAGCGCTGTCATTGGTACACCAATTG 402

DB

156 TCCTTCCTCGGTTACATCTCACCCTACTGCTCAAGCTCTCTCTGATACCGCTTTT 215

QY

403 CGAATCGC 410

DB

216 CCAACCCC 223

483 bp mRNA linear EST 20-DEC-2000

clone NF076F01EC 5', mRNA sequence.

BF646516

BF646516

BF646516.1

GI:11911646

barrel medic.

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 483)

Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,

Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation -

Center for Medicago Genomics Research

Unpublished (2000)

Contact: Dixon RA

Plant Biology Division

The Samuel Roberts Noble Foundation

| | | | |
|--|-----|---|-----|
| Db | 155 | TCGCTTCGGGTTACATCCTCACCGCTACTGGTCTCTCAAGCTCTCTCTGTGATACCGCTCTTT | 214 |
| QY | 403 | CGAATGCG | 410 |
| | | | |
| Db | 215 | CCAACCCC | 222 |
| RESULT 14 | | | |
| AI737351/C | | | |
| LOCUS | | | |
| DEFINITION | | | |
| AI737351 591 bp. mRNA linear EST 02-FEB-2000 | | | |
| 606039DD04.x1 606 - Ear tissue cDNA library from Schmidt lab Zea | | | |
| mays CDNA, mRNA sequence. | | | |
| ACCESSION | | | |
| AI737351 | | | |
| VERSION | | | |
| AI737351.1 GI:5058875 | | | |
| KEYWORDS | | | |
| EST. | | | |
| SOURCE | | | |
| Zea mays. | | | |
| Zea mays. | | | |
| ORGANISM | | | |

| REFERENCE AUTHORS TITLE JOURNAL COMMENT |
|--|
| <p>Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 591) Walbot, V. Maize ESTs from various cDNA libraries sequenced at Stanford University Unpublished (1999) Contact: Walbot V Department of Biological Sciences Stanford University</p> |

Stanford University
355 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Platform: 606039 row: D column: 04.
Location/Qualifiers
1. .591
/organism="Zea mays"
/cultivar="Obi-03"

```

/cultivar="Onio03"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt
lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/notes="Organ: Immature ear; Vector: pbk-CMV; Site_1: ECORI
; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
lab"
BASE COUNT      142 a    167 c    124 g    158 t
ORIGIN
Query Match      3.7%:  Score 36.6:  DB 9:  Length 591:

```

| | | | | | |
|---------|-----------------------|---|----------------|------------|-------------|
| | Query match | 57.8% | Score 36.0; | DB 9; | Length 391; |
| | Best Local Similarity | 50.3%; | Pred. No. 9.6; | | |
| Matches | : 90; Conservative | 0; | Mismatches | 89; Indels | 0; Gaps |
| QY | 629 | TTTGTGAGTCACATCACAGAGGACCCTTGGAAGATTCAAGTTCCAGGGTCTCGAGTCCGT | 688 | | |
| | | | | | |
| Db | 536 | TTTTCAGGGTGGAGTTCCCCGACTCGCTGGCACCAGAGCATGCAAGGCTCTCGAAGTCGT | 477 | | |
| | | | | | |
| QY | 689 | ACCTGTGCCTTAGCAAAAGCGTACACTCGATTGACATGGAGTTCATCTGTTTCAAGAATC | 748 | | |
| | | | | | |
| Db | 476 | ACTTTCGCCCAAAGCCTTCGTCCAAGCTGCACAGACATGGAGATGATGAATTCGAGGAGAC | 417 | | |
| | | | | | |

| | | | |
|------------|--|---|--------|
| Db | 416 | AACCATGCATGATCTGAACAACATCGAGGAAGAGATGCCGAGGAGCCTCATGCTG | 358 |
| RESULT 15 | | | |
| AW056085 | | | |
| LOCUS | AW056085 | 638 bp | linear |
| DEFINITION | 660004A04.y1 650 - Mixed stages of anther and pollen Zea mays cDNA, mRNA sequence. | EST 27-SEP-1999 | |
| ACCESSION | AW056085 | | |
| VERSION | AW056085.1 | GI:5928793 | |

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:30:47 ; Search time 2758 Seconds
(without alignments)
10562.693 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000
Perfect score: 1001
Sequence: 1 aactgtcttttcacctcg.....tccgtcaaggctggtttg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|--------------------|
| 1 | 408.6 | 40.8 | 10934 | 8 | FSESYNIA | Z18755 Fusarium sc |
| 2 | 405.4 | 40.5 | 2949 | 8 | FSESYN | Z48743 Gibberella |
| 3 | 282.6 | 28.2 | 9633 | 6 | BD013055 | BD013055 Cyclic de |
| 4 | 282.6 | 28.2 | 9633 | 23 | BD010088 | BD010088 Cyclic de |
| 5 | 240.2 | 24.0 | 46899 | 6 | A36768 | A36768 Sequence 1 |
| 6 | 240.2 | 24.0 | 46899 | 6 | AR050554 | AR050554 Sequence |
| 7 | 240.2 | 24.0 | 46899 | 8 | TISACYS | Z28383 T.niveum (A |
| 8 | 203.2 | 20.3 | 1497 | 8 | FPA345016 | AJ345016 Fusarium |
| 9 | 189.2 | 18.9 | 1713 | 6 | A36771 | A36771 Sequence 4 |
| 10 | 189.2 | 18.9 | 1713 | 6 | AR050556 | AR050556 Sequence |
| 11 | 157.2 | 15.7 | 3155 | 8 | COPEPSYNH | X96559 C.Oligospor |
| 12 | 71.6 | 7.2 | 10513 | 1 | AF023464 | AF023464 Bacillus |
| 13 | 66.8 | 6.7 | 7681 | 1 | AF023465 | AF023465 Bacillus |
| 14 | 66.8 | 6.7 | 8371 | 1 | AF087452 | AF087452 Bacillus |
| 15 | 59 | 5.9 | 39822 | 1 | BSPEPSYN | Z34883 B.subtilis |
| c 16 | 59 | 5.9 | 233780 | 1 | BSUB0010 | Z99113 Bacillus su |
| 17 | 58.8 | 5.9 | 56917 | 1 | AME16952 | Y16952 Amycolatops |
| c 18 | 54 | 5.4 | 18492 | 1 | AE004666 | AE004666 Pseudomon |
| c 19 | 53 | 5.3 | 22444 | 1 | AE004669 | AE004669 Pseudomon |
| 20 | 51.6 | 5.2 | 37200 | 1 | SCF53 | AL035640 Streptomy |
| 21 | 51.2 | 5.1 | 107379 | 1 | SHGCP1R | X86780 S.hygrosco |
| 22 | 51 | 5.1 | 8992 | 1 | SV117268 | Y17268 Streptomyce |
| 23 | 50.4 | 5.0 | 40989 | 1 | AF204805 | AF204805 Nostoc sp |
| 24 | 50 | 5.0 | 41599 | 1 | AB050629 | AB050629 Bacillus |
| 25 | 49.6 | 5.0 | 1200 | 6 | AR170980 | AR170980 Sequence |
| 26 | 49.6 | 5.0 | 1680 | 1 | AF172065 | AF172065 Streptomy |
| 27 | 49.6 | 5.0 | 9389 | 1 | STE250581 | AJ250581 Streptomy |
| 28 | 49.6 | 5.0 | 32748 | 1 | AB070951 | AB070951 Streptomy |
| 29 | 49.2 | 4.9 | 7178 | 1 | MXA6977 | AJ006977 Myxococcu |
| c 30 | 49 | 4.9 | 10569 | 1 | AE009298 | AE009298 Agrobacte |
| 31 | 49 | 4.9 | 31220 | 1 | AE008316 | AE008316 Agrobacte |
| 32 | 48.8 | 4.9 | 17277 | 1 | AF237701 | AF237701 Pseudomon |
| 33 | 48.2 | 4.8 | 37856 | 6 | AX024319 | AX024319 Sequence |
| 34 | 48.2 | 4.8 | 37856 | 6 | AX024212 | AX024212 Sequence |
| 35 | 48.2 | 4.8 | 40897 | 1 | AF004835 | AF004835 Brevibaci |
| c 36 | 48.2 | 4.8 | 333500 | 1 | AP003590 | AP003590 Nostoc sp |
| c 37 | 48 | 4.8 | 77457 | 1 | AF210249 | AF210249 Streptomy |
| 38 | 47.2 | 4.7 | 26477 | 1 | SCE29 | AL035707 Streptomy |
| 39 | 46.8 | 4.7 | 6465 | 6 | AX433751 | AX433751 Sequence |
| 40 | 46.8 | 4.7 | 8268 | 6 | AX433754 | AX433754 Sequence |
| 41 | 46.8 | 4.7 | 14452 | 1 | AB070955 | AB070955 Streptomy |
| 42 | 46.6 | 4.7 | 4918 | 1 | AF121767 | AF121767 Bacillus |
| 43 | 46.6 | 4.7 | 28798 | 1 | U95370 | U95370 Bacillus li |
| 44 | 46.6 | 4.7 | 32386 | 1 | BLAJ5061 | AJ005061 Bacillus |
| 45 | 46.4 | 4.6 | 39436 | 1 | ASP269505 | AJ269505 Anabaena |

ALIGNMENTS

RESULT 1
FSESYNIA
LOCUS 10934 bp DNA linear PLN 11-SEP-2001
DEFINITION Fusarium scirpi esynl gene for enniatin synthetase.
ACCESSION Z18755
VERSION Z18755.3 GI:15591913
KEYWORDS enniatin synthetase; esynl gene.
SOURCE Fusarium equiseti.
ORGANISM Fusarium equiseti
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE 1 (bases 1 to 10934)
AUTHORS Haese,A., Schubert,M., Herrmann,M. and Zocher,R.
TITLE Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldeipeptide formation


```
QY 565 GACCCCTTCAGCAGCTCAAAACCGTCGTATGCGCATCGAAGTCCGCGAGAGGCTTCGGT 624
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8324 GCCCTCTCCAGAAGCTGCAGCGCCGCTGTCGCGCTCTCAAGTCCGCGAGAGCTCCAGA 8383
QY 625 CTTTACTTCCATCGTACATGATCCCATCGAAGATGTTGTTCTGACAGAGATGCCCTCTCA 584
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8384 CGCTGGTCCCGCTTACATGTTCTCCGAATATCGTGTGCTGACACGATGCCCTCTCA 8443
QY 685 AGCCCAATGGTAAAGTTGACCGGAAGAACTCTCTCGCAGGCGAAGGTTGTACCGAAGC 744
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8444 ATACTAAGCGCAAGATCGACAGAAGAGCTTACCGGTAGACGACGACACTGCCGAGC 8503
QY 745 AGCAGACAGCAGCGCGTTTACCGACATTTCCCATCAGTGAGTGCAGATCTTCTTTGCG 804
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8504 AGCAGACTGCGCGCGCTGTCGCGGACTTCCCTATCTCTGATCGAGATCAGCGTGTGCG 8563
QY 805 RAGAAGCCACATGAGTGTGTCATGAAGGTGACATTACCGATCATTCTTCATCTCG 864
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8564 AGGAGGCAACTGAGGCTCTTGGAAATGAAGGTGAAATCAGCGATCATTCTTCCAGCTCG 8623
QY 865 GTGGACACTCTCTTGTGGCCAGGAGCTCATTCTCGTATGACCAAGCACTCAAGGTCC 924
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8624 GGGGTCACTCTCTCTGCTAGAACTCATTCTCGCATCCAGCACCGCTCCATGTGC 8683
QY 925 GTATCAGCTGTCAAGATGCTCTTTGACCATCTCTGATTTGCGGATCTAGCATCTGTCTCC 984
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8684 GGGTACTGTGAAGACGCTATTTCGACAGCCCTGCTCTTGGCGATCTGCGCAGTCAATCC 8743
QY 985 GTCAAGGCGTGGTTTG 1001
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8744 GTCAAGGACTTGCTATG 8760
```

```
RESULT 2
FSENSYN 2949 bp DNA linear PLN 20-FEB-2002
LOCUS Gibberella pulicaris partial esynl gene for enniatin synthetase.
DEFINITION Z48743
ACCESSION Z48743.2 GI:16444855
VERSION enniatin synthetase; esynl gene.
KEYWORDS Gibberella pulicaris.
SOURCE Gibberella pulicaris
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Nectriaceae; Gibberella.
REFERENCE 1
AUTHORS Burmester,J., Haese,A. and Zocher,R.
TITLE Highly conserved N-methyltransferases as an integral part of
JOURNAL peptide synthetases
MEDLINE Biochem. Mol. Biol. Int. 37 (2), 201-207 (1995)
PUBMED 96113556
8673002
REFERENCE 2
AUTHORS Glinoski,M., Hornbogen,T., Haese,A., Doller,A. and Zocher,R.
TITLE Nonribosomal Biosynthesis of N-methylated peptides in fungi
JOURNAL Unpublished
AUTHORS Burmester,J.
TITLE Direct Submision
JOURNAL Submitted (17-MAR-1995) Burmester J., Universitaet Zuerich,
AUTHORS Biochemisches Institut, Winterthurer Str. 190, Zuerich,
TITLE Switzerland, CH - 8057
JOURNAL On Oct 25, 2001 this sequence version replaced gi:732698.
COMMENT Location/Qualifiers
FEATURES
source
1. .2949
/organism="Gibberella pulicaris"
/strain="BBA 63933"
/isolate="wild type"
/db_xref="taxon:5128"
/clone="pCM2"
/clone_lib="1.5 kb and 2.9 kb PCR fragments in
pBluescript"
1. .2949
gene
```

```
/gene="esynl"
<1..2949
/gene="esynl"
/function="amino acid activation, N-methyltransferase"
/codon_start=1
/product="enniatin synthetase"
/protein_id="CAA8634.2"
/db_xref="GI:16444856"
/db_xref="SPTREMBL:Q00868"
/translation="EFNTGILKASLAIPLDVRSPVARMKIDILSSVSGNTIVIMGTGV
EDPGGLPQLELVRLTDTDFETIEDVONMSRPSATSLAVVFTSGTSGKPKGVMLTEHR
AIYRLVKSDFNPFSPARMVFNAAEDGASWEMFWMLNGTGYVCYDIYLAITLQDKE
LAAYFAKRVNCAFLAPAMLKLYLDARELANLDFLAVGGKFPDRADAEKNTLVRG
NANYVGTETAGMISTCYTIPKDEAFITGVQVLSYNGAYVMDPNQOLAGLVYNGE
HLFGDVGRTYKPELNKRNREIDVTIEKTRAYGTGDMRMRVGLLEFFGRDLNQ
FKMRQRTAGEVESAMLSHKLNNALVLRGQEGEOLQEMVGFVIAADDNDDEEE
TCNOVEGMDHFEESGMSDI STAVDQSAIGNDFKGTSMYDGNIDKGMQWMLDDAI
HTLNGQVPHHVLEIGTSGMILFNLNFGLOSQVGLDPSKSAVEFNRAIESSPFAG
KAKVHGNATDVKLGLPDLVFNVSQVYFPTPEILTEVIDGLLIAIPSAKRIFLGD
IRSYATNRHFLAARAIHTLPTNNATKDRVKQVQVELEDEESLVEPFAFTLKERR
PDVVKHVEIPKMKATNELIAYRTAVVHLRDEDEPYPTKESWIDFEKQMDKK
ALLDLRLSKDMSVAVSNITAYHTAFERRIVESLDEDSKDDAKDTLGGAAWLSAVRS
ETESRASLTVPLEFELANEGAFRVEVSAAWSONGALDAVFHFPSSNTDRTLQFP
TDNQLRSSLTLANRPLQKLRRAALQVRESLSQSLVPTVMVPTNVLVLDMLPTNGK
IDRKELTRARTLPKQTAAPVDFPISDIEITLCEBEATEVEGKMEVSDHFFQLGGH
SLLATKLISRIQHLRVTVKDFDPSVFDLAVIIHROGLAMQNPVABGQDKQGWSS
RVAPRTEVEKM"
BASE COUNT 701 a 838 c 767 g 643 t
ORIGIN
Query Match 40.5%; Score 405.4; DB 8; Length 2949;
Best Local Similarity 65.3%; Pred. No. 1.4e-112;
Matches 677; Conservative 0; Mismatches 321; Indels 39; Gaps 4;
QY 1 AACCTGCTTTCTTCACTCGTTGAAGACAGGTTTCCAGTCTGTGGTGAACATGTTGAGA 60
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1838 AGCCCTGCATCTTCAGACTCTGAAGACGAGGACCGGATGTTGTCAAGACGTTGAGG 1897
QY 61 TCCTGCCAAAGAACATGGAAGCTGTGAATGAGTCAAGTCGCTATCGATATGCCGTGTTG 120
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1898 TCATCCCCAAGAACATGAAGGCCACCAATCATCGCTACCGCTACACAGCTGTTG 1957
QY 121 TCACGTTGCGGGTTCACITTTGGAGATGAGCTTGTGCTTCCGGTTGAGAAGATGACTGGA 180
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1958 TGCATCTGCGGGATGAA---ACGGAGGAGCGCTGTATACCCCATGAAAAGGATGAGTTGA 2014
QY 181 TCGACTTTCAAGCGAATCAATTGAACAGAGATC---ACTGGTGACCTTCTCAAGCTCTT 237
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2015 TTGACTTTGAGGAGAAGCAGATGGATGAAGAGCTCTTCTTGATCACTCGCGCTCTCCA 2074
QY 238 CAGATGCTGCTATCATGCGCAGTCAGCAAAATTCCTTTTCGAAATTCACGGCTTTTGAAGAC 297
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2075 AGGATGCTATGAGCGTGGCTGTTAGCAACATCACCTACGCCACCTGCTTTGAGCGCAG 2134
QY 298 AGGTGCTGCTTCCCTC-----AATAGCAACATCGATG 330
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2135 GCATCGTTGAGTCTCTCGACGAGGATPAGCAAGGATGACGCCAAGGATACACTCGCGCGTG 2194
QY 331 AGTGGCAGCTATCAACCATTCGTCAGCGCGGCGGACGCTCATCATCTATCGTTCCCG 390
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2195 CAGCATGGCTCTCAGCAGTTCGTCGAAACCGAGAGCGTGCCTCACTACTGTTCCCG 2254
QY 391 ACATCTTTTCGCTATGCTGGGGAAGCGCGGTTCCGCTGTCAGAGTTCAGTTCTGCAAGACAGT 450
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2255 AGCTCTTCGAGATCGCCAACAGGCTGTTTCCGTGTCGAAGTTCAGCGCGCTCGCCAGT 2314
QY 451 GGTCTCAGAAATGGTCACTGGAGCGTGTTCATCA-----TTGTTGTTCCCAAGGCG 504
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2315 GGTCTCAGAAACGGTGTGTCGACGAGTCTTCCACCACCTTCCCATCTCCCAACACCGACG 2374
QY 505 GTACTCTGTGCAACTTTCTCAGGACCATCACTTCGAGGGTCTGTGATCTCTCCCTCACCACATC 564
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2375 GCATCTCATCAGTTCCCTACGGATTAACAGCTTCAGCTACCTACCTCCCTCGCCATC 2434
```


[illegible]

| FEATURES | | Location/Qualifiers | |
|----------|--|---|----------|
| source | | | |
| gene | | /organism="Tolypocladium inflatum" | 1..46999 |
| | | /strain="ATCC34921" | |
| | | /db_xref="taxon:29910" | |
| | | 885..46730 | |
| | | /gene="sima" | |
| | | 885..46730 | |
| | | /gene="sima" | |
| | | /codon_start=1 | |
| | | /evidence=experimental | |
| | | /product="cyclosporine synthetase" | |
| | | /protein_id="CAA82227.1" | |
| | | /db_xref="GI:440169" | |
| | | /db_xref="SPTREMBL:Q09164" | |
| | | /translation="MGAIGQDMAYDRLANPSRASSISSNRYSEPVQCSFAQGRMLFLH QKLGASNDITPAAIRLGRHGLDIDLNAASRALTORHETLRTFTFKEDQGVGVQVHLS GLERGLRTVDSRDLAQLAEETMKFLESEPPAMRVALLKVAEDHHLISLVHHII SDSRLOTIQOELGELYTAASQGSISACPLGPIQIYRDLTFTWQNOQDEQVABERQL GYWLEQDNNTPAELLTELPPAIPSGETGKI SFQIDGSHVHELLACFQSOQVTAAY LLAARVAFHFLTCGAEDATGAPVANDRPELENMVAPLATLOCMRVVLDEDDTFESV LRQINSWTEAHANDVPFERISALLPGSDTDSRHPVLQMLPALHPAQDTGRANWGF LEAETLQSAAPTREFDEMHLFEGEDDANDVLFSTGLFDAEIRSVSVIFREVRGRI SEPAHVKTMTPLDGLAAIRDMGLDITDTPREASVDMFQEQVALNPASFATVADA SSRISYSELDSKQLAAMRLRQLKETLIGVLSPECTETWYFGLIKLHAYLPL DINYLARIESILSADGHLVLGNSVQPKVDVPDVELIRISDALNGSOVNLGAK QATKPSATDLAYVIFTSGTSGCKEHSVTHBGRIVLRVLCGTNIIISPAQAAVPTAHLAN TAFDLSWEIYTPILNGTGLVCIEHSVTLDSKALEAVFTKEGIRVAFAPLJKOCLA DRPALFAGLDLAIADRFDRDLAHKSLVKHGVNAYGPTENSVSVTIYSVSEASP FVTGVPVGRATNSGAYVMDQDLVSPGMELVSGDGLARGYTDSDLDRKNRFVV QIDGESIRYISGTDARYSLKGGIIEFGMDQOVKIRGRIEPAIEPAHEALNLSQVR DAAVYIRQREEEEPAMIAFVYTGTLPLNPNKNGVDRKALQCSAKTVKSKLVSQVAFESLE RRVQMLLPVYMPAMIAVLDHLPNPNKNGVDRKALQCSAKTVKSKLVSQVAFESLE TEAVLCIEYRSLGVGVGVTIDNRFDLGSHLTAMKLAARISQRLDIOASVATVPEQPM LADLAATFORGSTLYSVITPTEYTGVPVEQFAOGRWFLQALNTGASVWMLTVRLR GHLVDALGTALLALEKHEFLRTTTFEERGQVGMQVHSSLMGELRLIDISEKSGTAA HEALMEOSTRFDLTREPGRVRAKLLADHHIFSIVMHHIVSDGSLDULLRHELQGL SAALRGQDPSLRLEPLIQYRDAFAWQDSQQAARQOLEYMTWKOLADPFAELLT DFRPSLTLSGKAGKVPVAIEGSLYDITLQVFSRTHQVTSFAVLEAARFAAHRFTLQSDN ATIGVPSANRRNPELENVIGFVNTQICRITIDENDNFESLRQVRSVTTAAQDNQDV PFEQVSSLMSPSSSRDASRNPVLQMLFALHQODLRFKIOLEGETEEVITPEETVRFDI BEHLQGNASKLSGDIIFAADLFEAETIRGVVSVPQEVLRGRLOQOPPTPIWMLPTLDGI PELEMMGLLHHVKTIDIPRNMVSVDVFOQVRLSAEATAVIDSSRMSYIELDQSDQV AAWLROLPAETFAVALPRSCAVIALFGILKAGHAYLPLDVPNVPAARLRAILAEV KGEKLVLLGAGEPSPGQSEVSVIRIADATSPAGHASLRDGSKPTASGLAYVIFTS GSTGPKGVMIHRGVLRKQTNILSLPAPOTFRMAHNSNLAFAASIEWEFTALLN GGLVCIDRFTILDAQALEALFRELHINIALFPALLKQCLTDAATAIKSLDLYVGG DRIDTADAALAKALVKSEVYNAYGPTENTWMLTYSTADTREFVNGVPVIGRAYVNSGV YVMOQOOLVPLGVNGLVVTGDLARGYTNPALSDBRFVDVYIARQQLIRAYTGDRA YRPRKDGVEFFGRMDHQVKVRGHRHIELAEVEHALLSSGADVAVVVSNQSDNQQVE MVAFTIADQNTLQEAASNNQVQWESHFTTAYADITADQNTLGRDFTSMVTSQDS TLIDKREMOEWLDMRTFLDQGAAGHVEITGTGTGMVLENLQAGLKSIGLEPSSQS QVAFVNAQOTFPGLEGKQAVHVGTAMDTRLRSALVSLDPLIVINSVQYPPSREYLAEV VEALVIRPGVRRIFPGDMRTVATHKDLVARAVHTNGSKVTRSKVQOEVARLELEEE LLVDPAPFTSLKESLSEIEHEVILPKMKVNNELSSYRYGAVLIRHNHNGSRSIH KINAESWIDFASQMDRQGLARLLKENKDAEIAVFNIPYSKTIIVERHAKSLADDDH GDDTHSSIDGVAWISAAREKASQCPSLDVLHDVLAEDGAFVEVSWARORSNGALD VFFHFQPTENESRALVDFTDYQQAARSLTRNRLQVRSESRRIEAQREQQLWLLPA YMIIRFIVYLNQMLPLTSGKYDRKELTIRAKVTAARTPSSSELVAPRDEIAETICKEFK DVLGVGVTIDNFNFVGGHSLATKLAARSLNOLNAQIAVFKDIFRPLVLAADTAATQQ DTEHNPTLPTSYTGPVEQSPAQRWFLDQNLNGATWYLMFPFARLPLGVVLSAALQ ALLALEERHETLRTTFIEQEGICMQVHPAPKELRVIDVSGSEESTIOKILEKQTT PFLNAESPGFRALLKTGDEDEHILSTVHHHAISSDWSVDIFQOEIQGFYSAILGRHDP LAQAPUSIQYRDEFATWORQIFQVAEHRRLQYMTKOLADNKPAAELLTDFKRPPLMUSG RAGEIPVVDGLIYKQDFCIRIQVTAFTVLLAAAFRAAHRMTGTEDATIGTPIANR NRPELEGLIGFFVNTQCMRITVDQSFETLVHVQTRFETLVLAHANQDVPPEQVSNIL PGSDTSRNPVLQMLFALHSONLKVRLGIEEIEIISIAETTRFDIEHLHQEAERL NCSIVYAADLFVETIQSVITFOGILQKGLGREGPEVMSMALDGLSLRSTGLLHP QQTDYPCDASVQIFKQOVANPDVIAVRESTRLSYADLRDSDQVACVWLSRGTAP ETFAVILAPASCEITIRITELDAKTDDINGLAASOPTAAISLAYVIFSGSTGRPKGVW TPTEPGMAEACETIRITETLDAKTDDINGLAASOPTAAISLAYVIFSGSTGRPKGVW EHRGTVRFTKQTNITSLKPSFHANLSNLFASDASVMEVFTLLNGTGLDIDYFTLL ESTALEKVFDOORNNVALLPALLKQCLDNPALVKTLSVLYIGGRDLDAASAAKARG | |
| gene | | /organism="Tolypocladium inflatum" | 1..46999 |
| | | /strain="ATCC34921" | |
| | | /db_xref="taxon:29910" | |
| | | 885..46730 | |
| | | /gene="sima" | |
| | | 885..46730 | |
| | | /gene="sima" | |
| | | /codon_start=1 | |
| | | /evidence=experimental | |
| | | /product="cyclosporine synthetase" | |
| | | /protein_id="CAA82227.1" | |
| | | /db_xref="GI:440169" | |
| | | /db_xref="SPTREMBL:Q09164" | |
| | | /translation="MGAIGQDMAYDRLANPSRASSISSNRYSEPVQCSFAQGRMLFLH QKLGASNDITPAAIRLGRHGLDIDLNAASRALTORHETLRTFTFKEDQGVGVQVHLS GLERGLRTVDSRDLAQLAEETMKFLESEPPAMRVALLKVAEDHHLISLVHHII SDSRLOTIQOELGELYTAASQGSISACPLGPIQIYRDLTFTWQNOQDEQVABERQL GYWLEQDNNTPAELLTELPPAIPSGETGKI SFQIDGSHVHELLACFQSOQVTAAY LLAARVAFHFLTCGAEDATGAPVANDRPELENMVAPLATLOCMRVVLDEDDTFESV LRQINSWTEAHANDVPFERISALLPGSDTDSRHPVLQMLPALHPAQDTGRANWGF LEAETLQSAAPTREFDEMHLFEGEDDANDVLFSTGLFDAEIRSVSVIFREVRGRI SEPAHVKTMTPLDGLAAIRDMGLDITDTPREASVDMFQEQVALNPASFATVADA SSRISYSELDSKQLAAMRLRQLKETLIGVLSPECTETWYFGLIKLHAYLPL DINYLARIESILSADGHLVLGNSVQPKVDVPDVELIRISDALNGSOVNLGAK QATKPSATDLAYVIFTSGTSGCKEHSVTHBGRIVLRVLCGTNIIISPAQAAVPTAHLAN TAFDLSWEIYTPILNGTGLVCIEHSVTLDSKALEAVFTKEGIRVAFAPLJKOCLA DRPALFAGLDLAIADRFDRDLAHKSLVKHGVNAYGPTENSVSVTIYSVSEASP FVTGVPVGRATNSGAYVMDQDLVSPGMELVSGDGLARGYTDSDLDRKNRFVV QIDGESIRYISGTDARYSLKGGIIEFGMDQOVKIRGRIEPAIEPAHEALNLSQVR DAAVYIRQREEEEPAMIAFVYTGTLPLNPNKNGVDRKALQCSAKTVKSKLVSQVAFESLE RRVQMLLPVYMPAMIAVLDHLPNPNKNGVDRKALQCSAKTVKSKLVSQVAFESLE TEAVLCIEYRSLGVGVGVTIDNRFDLGSHLTAMKLAARISQRLDIOASVATVPEQPM LADLAATFORGSTLYSVITPTEYTGVPVEQFAOGRWFLQALNTGASVWMLTVRLR GHLVDALGTALLALEKHEFLRTTTFEERGQVGMQVHSSLMGELRLIDISEKSGTAA HEALMEOSTRFDLTREPGRVRAKLLADHHIFSIVMHHIVSDGSLDULLRHELQGL SAALRGQDPSLRLEPLIQYRDAFAWQDSQQAARQOLEYMTWKOLADPFAELLT DFRPSLTLSGKAGKVPVAIEGSLYDITLQVFSRTHQVTSFAVLEAARFAAHRFTLQSDN ATIGVPSANRRNPELENVIGFVNTQICRITIDENDNFESLRQVRSVTTAAQDNQDV PFEQVSSLMSPSSSRDASRNPVLQMLFALHQODLRFKIOLEGETEEVITPEETVRFDI BEHLQGNASKLSGDIIFAADLFEAETIRGVVSVPQEVLRGRLOQOPPTPIWMLPTLDGI PELEMMGLLHHVKTIDIPRNMVSVDVFOQVRLSAEATAVIDSSRMSYIELDQSDQV AAWLROLPAETFAVALPRSCAVIALFGILKAGHAYLPLDVPNVPAARLRAILAEV KGEKLVLLGAGEPSPGQSEVSVIRIADATSPAGHASLRDGSKPTASGLAYVIFTS GSTGPKGVMIHRGVLRKQTNILSLPAPOTFRMAHNSNLAFAASIEWEFTALLN GGLVCIDRFTILDAQALEALFRELHINIALFPALLKQCLTDAATAIKSLDLYVGG DRIDTADAALAKALVKSEVYNAYGPTENTWMLTYSTADTREFVNGVPVIGRAYVNSGV YVMOQOOLVPLGVNGLVVTGDLARGYTNPALSDBRFVDVYIARQQLIRAYTGDRA YRPRKDGVEFFGRMDHQVKVRGHRHIELAEVEHALLSSGADVAVVVSNQSDNQQVE MVAFTIADQNTLQEAASNNQVQWESHFTTAYADITADQNTLGRDFTSMVTSQDS TLIDKREMOEWLDMRTFLDQGAAGHVEITGTGTGMVLENLQAGLKSIGLEPSSQS QVAFVNAQOTFPGLEGKQAVHVGTAMDTRLRSALVSLDPL | |

LVQTAQFNAYGPTENTVMSIYPIAEDPFFINGVPIGHAVSNSGAFVMDQNOQTTPPGA
MGLIVTGLGARGYTSSINTGRF INVDIGQVRAIYRGDRVRYRPKDLQIEFFGR
IDHQVKIRGRHIEPAEYEAIALSHDLVTDAAVTHVTHSQENDQLEQVFAVADVED
ESSNOVQOHPHDSIAYADITIDQOGLRDMSTMYDGLSIKKSOEWMLDDTM
RSLSDOPPGHLEVGTSIADWTLNREGGLOSVCGLPSPSATAFVNKAASFPGL
EDLRVEVGTATDIDRLDGLHAGLVVNSVAYFPQSDIYLAQLVRDLTKVPGVERIF
FDMRSHAINRDLKARVAHALGDKAKAELOREVRVRESEDELLVDPAAFTSLRTO
VENIKHVEILKRRMARTNELLSSYRAALVHLNDLAKPAHKVSPGAWAATKMDTA
LIRLRGTKISDHIAIANIENSKITVERTICESYDGLGDKADNSDRVLSAARNA
VKVASAIDLDIAQAGFVEISCAOWSNCALDAVPHHLGSPSSQSHVILIDFLT
DHGRPEALTNPJHRAOSRRVEROIRERLQTLTPAYMIPAQIMLVLDKELPLNANGKV
DRKOLTRAQTVPAKAOYAPAPARTIEIUVLCQESDVLGVLDIGIMENEFGLGCHSL
MAYKLAARISRLRTHSVKSEIFDHPRVCDLVLVQOQSAPHDPIYSTKTYGVPQSF
AQGLWPLQNFNGATWYLMPLVRLRAGMNVHALTAALLERHRELLRTTYEONG
VGKQNVNVTETRIIDLSNGDGLYTLTKKEQTAPFHELETPGRVALLRLGPGDY
ILSVMMHIIISDQVSVLQFQELGQFYSTAVKGLDPLSQTTPLHYRDFALKWQKPT
QESHEHQIYVQEOVLDSAPAEILLTOLPPRSILSGOAGEMSVTIEGALKNLEFCR
VHVTSFVLLAALAAHYRLTGSEDATIGTPIANRNRPELEQIIGFFVNTQIRIVY
NEDETFSVLQVRSATATAAFADQVFEKIVSTLLPGSDASRNPLQVLMFAVHOSK
NLGELKLENAHSEVPIETITREFLEPHLQDDKLEGSILYSTDFEAVSVQSLSV
FOELIRGLNGPDPVISTPLQDGLVQLRQGLDVQKTEYPRDSSYVDVHEQVSVN
PDSIALIHGSEKISYAQLDRSDRVARMLHRFSFSDTLTAVLAPRSCETIIAFLGIL
KANLAYLPDVKAPAAIDAIVSLPGNKILLCANVTPPKLOEAAIDFVPIDRFTT
LTDGTLQDGPTEPSSAQSLAYAMFTSGSRPKGVNMVHRNIVLKNVNVAKOPA
AARIAHISNAFASASHEIYAPLLNGGAIYCADYFTTIDFQALQETQOEHEIGAMP
PSSLKQCLVQAPDMISRLDILFAAGDRFSSVDALQAOQLVSGSVFNAYGPTENTILST
IYNVAENDSVNGVPIGSAVNSNGAYIMDKNQQLPAGVMELVVTDGLGARMDPK
LDADRFTQLVNSQVRAIYRGDRVRYRPKDFQIEFGRMDQOIKIRGRHIEPAEVE
QAFNDGCFQVDAVITIPENQOPEMAVFTAKDNSAREEATTOEGWEAHEFGGA
YANIEETESALGDFMGWTSYDGTEDIDKEMREWLNDTMRSLDGLKAGRVLEVT
GTGMIMNLSRSGOLERYIGLEPAASAEFVNNAKSPFGLAGRAEVHVGTAADVGTFL
QGLTSDMAVINSVAQYPTPEYIAETIKSLVQVPGMKRIYGLDMRSWMNRDFAAARA
AYSADNADRVORKNMEKEEELLVDPFAFTALASQLODRIQHVELIPKRMKAT
NELSYRYAAVLHISDEPLTYIKIDPEAMINFEGRSLTREALAQVLKENENASVATS
NIPKSVTVRHHIVRSUDODANAPESMDGSDWISAVRTAQQCHTPLASDLFDIAE
DAGRFVSVARHOSHODADVPHLKPATEDSRVLIKPTPHQCHPLKSLTNQPL
PAQSRRAELLIRGLQTLPLPYMIPSOITILDRMLNANGKVDRELRARAKTIOKSK
PVEDIVPNSRSEATVCKGTFVLGVEVITDNFNLGHSMLMATLAARLGRLOLNTIR
ISVRYDQPVADLAIQIORSAPHEIPKADYTPVPQSFQAGRLWFLDQNLVNGAT
WYLMPLGIRLHSLRVDALATAISALEORHEPLRTTTHEEDGVGVQVODHRPKDLRI
IDLSTQPKDAYLAVLHQEOTLFDLATEPCGRVALIRTHEEHEILSVMMHIIISDQV
VEVILDRMHREYSALRQDPMQIILPQYRDFAAWQTEQOVARHOFQOLWYH
LAUSTPAELLTDLPRSLGRANELPLTIEGRHDKLRAFRCVHQATPFIYLLAAR
AAHRLTGADATLGTPIANRPELNMIGTFVNTQCMRIAIEENDFESLVRVARS
TATSANQADVPFESIVSULLPGSDASRNPLQVILAVHSQDQLDLGLEGRDEAV
DSATSTREDFEHLFHADRLSGSVLYAKELFKLRTIESVSVFLETLRRLDQPLTP
LAVPLTDGVEITASKGLDVPRTDPRANIVEVFOQHVRATPDALVADKATSILTY
AQLDQSDRLAIWLSRRHMPETLVGLAPRSCETIIAMFGINKANLAYLPLDINSPA
ALRLSILSAVDGNKVLGLSGSVTAPQOENPEVEAVGQIETLACTGLDKTOGSNARSA
TSLAYLFTSGSGTKPKGVWVHRSVTRLAKPSNISKLPQGARVAHLANAFDASTW
EIAATLLUNGATLVCLDYHTVLDCTRKEVERESITVVTLPALLKOCVAAEIPETLAH
LDLLYTGDRVGHGDAMRARSVKIGMFSGYGTENTVITIEYEVDADEMFAVNGPTIG
KTVNSGAYVMDNRQOLVPSGVVGLVVTGDLARGYTDPSLKNRRIYITVNGESIR
AYRTGDRVRYRPHDLOTEFGRMDQOVKIRGRHIEPGEVESALLSHNSVODAAVICA
PADQDSAEVFAVAENTEDETOEEAVDQVQNETHEETAAYSEVKDLRQSEVGN
DFWGTSMYDGESEIDKIDMHEWLNDRMLDAREPGHVLEIGTGTGMWENLAKQPC
LQGVYGFPEPSAFVNDAAQSFPAKDGRSIVHVGTATDINKRAGIOPRLVYINSV
AQQFTPEYLFYRVAEALQYSPVERIVFGDMRTNAINRDSVSRALHTLEKANKRLV
ROMIYEEANEELLTDPAFTSLRTRLEKIKHVELILPKTRMKNELSKYRISYAAVLH
VRMSQESTTHOVSNWAIQFAADGLDROTLLINKLHKDAGTVAIGNIPYSKTIIVER
FVNKLSLEDDEMGQNSLDGSANVAARMAAQSCPSLDAMDVKELAQEAGYOVEVSHA
RQWSONGALDAIHFFHEPKEGARTLIEFTDYEGRNVNLTNRPLNSIOSRRLGTQI
REKLQTLPLPYMIPSRIMVLDQMPVNNNGKIDRKLVRRAIVAPKPRSAATRAPRNE
IEALDEFEDVLGTEIVSDNFDLGGHSLMATKLAARVSRRLDAHISIKVFDQDPV
LADLAASIQESAPHEIPQORTYFGAEQFQAQRLWFLDQLNLGATWYLMPLAIRIR
GOLVRAALSALFALRHEHRTLTTFEESGVGVQIVGEARNSDLRVDHVSITGDGGEY
LEVLRROOTVPDLSSEPGRWVCLVKTGEEDHVLISVMMHIIYDGHVSDDLRLGELQF
YSALRQDQDPLHANPLIOYRDEAANOREAKQVEHORQOLYWSQOLVDSFAELLT
DLRPSILNSGRAGSDVTIEGVSYGALQSCFRTRSVTTFVVLITVRIAHFRUTAVDD
ATIGTPIANRPELETIEGCFVNTQCMRISIAADDNFEGVLQVRNRTAAYANQDV
PFRIVPGLVGSVSRNPLVQLMFAVQSVEDYDQVRLEGLSVMPGEASTRFOME
FHLVPGDKQILSVLSDLFQEQITQNFVDFEQELRSVLDQPLTPISVLPFSNAIS
NLESILDLEMTSDYPRDRITVVDLFRQAAICPDSITAVKDSQSLTYAQLDEQSDRVA

AWLHERHMPAESLVGLVSPRSCETIIAIFGIMKANLAYLPLDVTAPDARLAAALDITVE
GERLLGAGVPGPQIQIIRLSTAYIAELSHATVDTVTSIPQPSASTSLAYVFTSGS
TGPKPGVIMHGRIVLRVTRNVNFPESGALPVSHFNLAADAAWTAETVAYLVNGG
TVPCIDRDTMLDIAALNSTKRKNRAAFTPAFKOCLAETPELVANLETLHTAGDR
LDGCDANLACKTAGGELFNLVGHPTENTAYSTFFVYVGEETFYVPGVGRGSLNSHAYI
IDRHQKLVAGVAGGELITGDGVARTYDSALNKDRFVYIDINKSGTSYKRTGOKAR
RFDGQOIEFGRMDQOVKIRGVRIPEGEVELTLLDHKSVLAAATVVRPPNPDPEMIA
FITIDAEDDVQTHKAIYKHLQILPAYMIPSHLVILQDMPVTDNGKDRKDLARAO
PQKRSTAAARVDDPEAVEAVLCEYSNLLVEVITDGFDPGLGSLALLAKLAARLSR
QLNTRSVKQVDEQDIIADLADIIRGSHRHPDIPATPYTGPQESQAQGRFLWFLLEOL
NLGASWYLMFPAIRMRGPLOTKALAVNALVHRHEALRTTFEDHDGVGVQVIOQPSK
QDUKILDSAVDDDTAYLAAKREQTAFDLTSEPGRWVSLRLGDDDDYILISIVMHHI
ISDGWTVLDVRELQGOFSAAIRGOEPLSQASLPQIYRDFVAVMROENQIKQAKOL
KYWSQOALRTPECFETDLRPSILSGEADAVPMVIGDVTYQLTDFCRHTQTSFVS
LLAARFTHRLTGTLDATVGTPIANRNRPELEGLIFFVNTQCMRAISQETFEESL
VQOVRVLTAFANQDVPFQIIVSTLLPGSDRSRNLVOVMFALQSQODQLRQLEG
MTDEALTEPLSLDLLEVHLFQEVGLSGSLIYSTDLFEVETIRGIIVLELILRRGL
EQPKRQRLMAMPIDGTITKLRDQGLLTVAKPAYPRESSVIDLFRQVAAAPATAVWDS
SSLTLYADLDQGNKLAHMLCQRMNAPETIIVAPRSCLTIVAFGLVKANLAYLPL
DVNAPARIETAILSAVPGHKLVLVQAHGPELGMTADTELQVIDEALSSGQHEQI
HASGPTATSLAYVMFTSGSTGPKGVMDHRSITRLVNSDVVATLTPPYRMANVSNL
AFDISVOEITALLNGTLCVLDYLLTLLDSKILYNFVEAQVNAAMFTPVLLKQCLGN
MPAIIISRLSVLENNVGRDLDAHDAVAASLIQODAVYNAYGPTENGWQSTMYKVDVNEP
VNGVPIGRSITNSGAYVMDGNQOLVSGVMEIVVYTGGLARGYTDLSALDEDFRVHVT
IDGENTAKITRGDRVIRKPDIEFFGRMDQOVKIRGRHIEPAEVEHALLGHDLVH
DAAVYLRKPNQPEMTAFITSQEDETIEQHSNKQVGGHEFDSVRYADIKDLDS
TFGHDFLGWTSMYDGVDPVYNEMKEMWDETASLLDNRPFGHILEIGSLDKNEIOELV
KVDGLQRYFTLDPAPSAAFVNEAVKPSLQAGKARVLGTALDGLSDKNEIOELV
VINSVAQYFTPSEVLKVKRAVVEVPSVKRFGDIFRQALNRDFLAARVAGLDNA
SKQIRKIAELSESEELLVDPAPFVLSRQPLNKHVEVLKLMKATNELSYRYA
AVLHISINESEQLLIQIDIPITAMVDFAAOTDSQGLRNLQOQGRDDVIAVGNIPYSK

Query Match 24.08; Score 240.2; DB 8; Length 46899;
Best Local Similarity 55.08; Pred. No. 7.5e-62;
Matches 563; Conservative 0; Mismatches 418; Indels 42; Gaps 3;
Qy 1 AACCTGCTTCTTCACCTCGTTTGAAGACAGAGTTTCCACGCTCTCGTGGAGAACATGTTGAGA 60
Db 21076 ACCTCGATCTTTTACATCTTTTGGCTACGCGCTTGGGTGAGAGATCAAGCAGCTCGAAA 21135
Qy 61 TCCTGCCAAAGAACATGGAAGCTGTGAATGAGCTCAGTGGCTATCGATATCGCATATCCCGCTGTG 120
Db 21136 TTTCCCCAGAGACCATCAAGCTACCAACAGACTCAGCAAGTACCGATATCCCGCAGTAC 21195
Qy 121 TGCACGTTCCGGGGTTCACCTTGGAGATGAGCTTGTGCTTCCGGTTGAGAAAGATGACTGGA 180
Db 21196 TACATGTGCTGGCTCGAGAGAACAACTCACTATACACCAAGTCTCTCCCAACGCTTGA 21255
Qy 181 TCGACTTTCAAGCGAATCAATTGNACAGAGACTCACTGGGTGAGCTTCTCAAGTCTTCAG 240
Db 21256 TAGACTTTGGCGGACAGCGGTCTCGACCGGACGACCTTCATCACTTCTGAAGGAGACA 21315
Qy 241 ATGCTGTATCATGGCAGTCAGCAAAAATTCCTTTTCAAAATCACGGCTTTTGAAGACAGG 300
Db 21316 AGNATGCGGGGACCGTCTCGTATCGGTAATATCCCGTACAGCAAGACCATGTTGAGCGGT 21375
Qy 301 TCGTCTGCTTCCCTCAATAGCAACATCGATGAGTGGCA----- 337
Db 21376 TTGTCAACAAGTCACTCAGCGAGGATCATATGAGGAAGGCGCAACACTCACTGGACGGAT 21435
Qy 338 -----GCTATCAACCATTCGGTCCAGCGCGGAGGGGACTCATCACTATCCGTTCCCG 390
Db 21436 CAGCTTGGGTTGACCGCGTCCGGATGCGCGCTCAAGCTGCCCATCGATCAATG 21495
Qy 391 ACATCTTTCGATTGCTGGGGAAGCGCGGTTCCGCTGTGAGAGTCAGTTCCTGACGACAGT 450
Db 21496 ATGTCAGGAGATGCTCAGAGGCGGGATACACAGGTTCGAAGTCAGTTCGGCGCTCAAT 21555
Qy 451 GGTCTCAGATGTTGACATGAGCGCTGTTTTCACATATGTTGCTCCCAAGG----- 503
Db 21556 GGTCCCAAGATGTTGCGCTCGATGCCATCTTCCATCATCTTCCAAACCGCCCAAGGAGGTG 21615
Qy 504 --CGTACTCTGCTCAACTTTTCCCTACGCGACCATCACTTCGAGGGGTCTGATCTCCTCACC 561


```
Db 21616 CTCGACACTTATTGAGTTCGCCGAGGATTACGAGGCCGGAATGTGAACACCTTAAACGA 21675
QY 562 ATCGACCCCTTCAGGAGCTGCAAAACCGTGTATCGCCATCGAAGTCCGCGAGAGGCTTC 621
Db 21676 ACCGTCCCTCGAAGAGCTTCAAGACCCCGCTTCCTGGACGAGATCCGCGAAGAGCTGC 21735
QY 622 GGTCTTACTTCCATCGTACATGATCCCATCGAAGATCGTTTCTGGACAAGATGCGCTC 681
Db 21736 AGACCTCTCGCGCTTACATGATCCCATCGCGCATGTGCTCTTGCATCAGATGCGCTG 21795
QY 682 TCAAGCCCAATGGTAAAGTTGACCGAAGGAACTCTCTCGCAGGCAAGGTTGTACCGCA 741
Db 21796 TCAACAACAAAGGCAAGATTACCGCAAGGAGCTTGTGCGGAGAGCTATCGTGGCCCGCA 21855
QY 742 AGCAGCAGACGACGCGCGTACCGACATTTCCCATCAGTCAAGTTCGAAAGTCAATCTTT 801
Db 21856 AGCCAAAGTCAGGCGTAC---TCGGGTAGCCCCCGCAAGTGAATGATCGAGCTATCTGA 21912
QY 802 GCGAAGAAGCCACTGAGGTGTTGGCATGAAGTTGACATACCGATCACTTCTTCAATC 861
Db 21913 GAGAGCAATTCGAGGACGCTGTCGGAACAGAGTCAAGCTGCTGGATAACTTCTTTGATC 21972
QY 862 TCGGTGGACACTCTCTTGGCCAGCAAGCTCAATTTCTCGTATCGACCAAGCACTCAAG 921
Db 21973 TCGGCGGCACTCACTTATGGCCAGCAAGCTGCGCCCGCGTTCAGCCGCGCTGTATG 22032
QY 922 TCCGTGACACTGTCAAGGATGCTTTGACCACTCTGTATTTGCGGATAGTACATCTGTCA 981
Db 22033 CCCATATTCCATCAAGATGCTTTGATCAGCGGCTGCTGCGGATCTTGGCGGTCCA 22092
QY 982 TCC 984
Db 22093 TCC 22095
```

RESULT 8

```
FPA345016
LOCUS FPA345016 1497 bp DNA linear PLN 14-SEP-2001
DEFINITION Fusarium pallidoroseum partial gene for enniatin synthase.
ACCESSION AJ345016
VERSION AJ345016.1 GI:15626354
KEYWORDS enniatin synthase.
SOURCE Fusarium pallidoroseum.
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE 1 (bases 1 to 1497)
AUTHORS Glinksi,M., Hornbogen,T., Haese,A., Doller,A. and Zocher,R.
TITLE Nonribosomal Biosynthesis of N-methylated peptides in fungi
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1497)
AUTHORS Zocher,R.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2001) Zocher R., Fakultät II, Chemisches
Institut, Franklinstr. 29, 10587 Berlin, GERMANY
COMMENT orf based on homology with Z18755 and Z48743.
FEATURES
Location/Qualifiers
1..1497
/organism="Fusarium pallidoroseum"
/strain="ATCC 74289"
/db_xref="taxon:100904"
/clone="pPallbre"
<1..>1497
/function="N-methyltransferase"
/note="orf"
/codon_start=1
/product="enniatin synthase"
/protein_id="CAC69934.1"
/db_xref="GI:15626355"
/translation="QFKMRKRIESAEEVAAMLSHVSVRDAVVVKDDKADLVGF
VVIDHDSLEGDANDNQEGQDHFETEMYADIGDINFTIGKDFKWTSHYDGEID
KVMEQEWLEDDTIKTURDQAFQHVLEVTGSGMILFLNLDGLQSTRGLEPSKSAAF
```

CDS

```
RESULT 9
LOCUS A36771
DEFINITION Sequence 4 from Patent EP0578616.
ACCESSION A36771
VERSION A36771.1
KEYWORDS GI:2294037
SOURCE Neocosmospora vasinfecta.
ORGANISM Neocosmospora vasinfecta
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreales; Nectriaceae; Neocosmospora.
A36771 1713 bp DNA linear PAT 05-MAR-1997
Sequence 4 from Patent EP0578616.
```

```
BASE COUNT 329 a 407 c 413 g 348 t
ORIGIN
Query Match 20.3%; Score 203.2; DB 8; Length 1497;
Best Local Similarity 61.1%; Pred. No. 1.2e-50;
Matches 413; Conservative 0; Mismatches 233; Indels 30; Gaps 4;
QY 1 AACCTGCTTTCTTCCACCTCGTTGAAGACAGGTTTCCAGGCTGTGTTGGAACATGTTGAGA 60
Db 824 AGCCAGCTTTCTTCCAGGCACTACAAGATCGCTTTCCCGACTTGTGCCACCATGTTGAGA 883
QY 61 TCCTGCCAAGACATGAAGCTGTGAATGAGCTCAGTCCGCTATCGATATGCCGCTGTG 120
Db 884 TTCGCCAAGAACATGCTGCTACAAATGAGCTTAGTCCGCTATCGGATGCGGCTGTG 943
QY 121 TGCAGCTTCGGGGTTCACCTTGGAGATGAGCTTGTGCTTCGCTTCGAAAGATGACTGA 180
Db 944 TGCATATAGCCCAT---CACGACTCGTCCCTGTGCATACGATTGAGAAGGTGCTTGG 1000
QY 181 TCGACTTTCAAGCGAATCAATTGAACAGAACTCACTGGGTGACCTTCTCAAGTCTTCAG 240
Db 1001 TCGACTTTGGGCGCATCAGCAATGATCGTAATCTCTCTTACAGTTTCTACGACGCTCA 1060
QY 241 ATG---CTGCTATCATGGCAGTCAGCAAAATTCCTTTCGAATACAGGCTTTGAAGAC 297
Db 1061 AGGGCTCATACGCGTGGCCATAAGCAACATACCTTCGCCAAGACCCGCTCTTTGAGCG 1120
QY 298 AGGTGCTCGCTTCCTCCATAA-----TAGCAACATCGATGATGGCAGCTAT 342
Db 1121 AGATCGTTGAGTCCCTTCAGGAAGAGGATAAGTCCAACCTCAGCGCGCGCTTGATCT 1180
QY 343 CAACCATTCGGTCCAGCGCGGAGGCGACTCATCACTATCCCTCCCGACATCTTTTCGCA 402
Db 1181 CAACCATTCGGTCCAGCGCGGAGGCGACTCATCACTATCCCTCCCGACATCTTTTCGCA 1240
QY 403 TTGCTGGGAAGCCGGGTTCGCTGCGAGGTCAGTCTTCGACGACAGTGGTCTCAGATG 462
Db 1241 TCGCTCAAGACGCTGGGTTCGCTGAGGTCAGCGCGCGCTCAGTGGTCTCAGAGCG 1300
QY 463 GTGCTATTCGACGCTGTTTCCATCATTTGTGCTCC-----AAGGCGCTACTCTGG 513
Db 1301 GTGCGCTTGTATCGCGCTTTCACCACTTCCCGTCCCTTCGCGACACTCGTCGACGCTGA 1360
QY 514 TCAACTTTCTTACGAGCAATCACTTCGAGGCTGTGATCTCTCACCAGATCGACCCCTTC 573
Db 1361 TCAAGTTCCGACCGGATAATCACTCCGAAGCTCAGCGACCCCTTGCAGTTCGCGCTAC 1420
QY 574 AGCGACTGCAAAACCGTGTATCGCATCGAAGTCCGCGAGAGGCTTCGCTCTTACTTC 633
Db 1421 AGGGATTACAGCGCGCTGTCGCGGCTTACAAAGTCGCGAGGATTCAGACTCTCTACTCC 1480
QY 634 CATCGTACATGATGCC 649
Db 1481 CCACCTACATGATGCC 1496
```

REFERENCE 1 (bases 1 to 1713)
AUTHORS Leitner, E., Schneider, E., Schoergendorfer, K. and Weber, G.
TITLE Cyclosporin synthetase
JOURNAL Patent: EP 0578616-A 4 12-JAN-1994;
SANDOZ LTD (CH)
COMMENT Other publication JP 6225773 940816
Other publication AT 140392 940515
Other publication AT 43793 940415
Other publication AT 398578 941227
Other publication AT 398434 941227.
FEATURES
Location/Qualifiers
1. 1713
source /organism="Neocosmospora vasinfecta"
/db_xref="taxon:40622"
BASE COUNT 389 a 491 c 469 g 364 t
ORIGIN
Query Match 18.9%; Score 189.2; DB 6; Length 1713;
Best Local Similarity 59.5%; Pred. No. 2.3e-46;
Matches 396; Conservative 0; Mismatches 258; Indels 12; Gaps 4;
QY 338 GCTATCAACCATTCGGTCCAGCGCGGAGGCGACTCATCATATCCGTTCCCGACATCTT 397
DB 992 GATATCGCCACTCAATCACGGGGAAGGAATGCCCTGCTCTCTCAGTGGCGGACCTGAT 1051
QY 398 TCGCATTTGCTGGGAAGCGGGTTCCTGTCGAGGTTCAGTTCGACGACGAGTGGTCTCA 457
DB 1052 TGAGATTGGTAAGGGATCGGCTTCCAAAGTTGAGACCAGCTGGGCTCGACACACTCCCA 1111
QY 458 GAATGGTGCAATGGACGCTGTTTCCATCATTTGTTGCTCCCA-----GGCGGTAC 508
DB 1112 GCGCGGCGGACTCGATGCTGTTTCCACCGATTTCGAAACCAAGACACTCGGGTCAATG 1171
QY 509 TCTGGTCAACTTCTACGACCATCACTTCGAGGGTCTGATCTC-CTCACCAATCGAC 567
DB 1172 CATGTTTCAGGTTCCCAACTGAACACAGGGGCGGCTTTCGACGAGTCTCAAGATCGCC 1231
QY 568 CCCTTCAGCACTGCAAAACCGTCGATCGCCATCGAAGTCCGGAGAGGCTTCGGTCT 627
DB 1232 CGCTACACCTTGGTTCAGAGCGCGGCTGGAGGCAAGGTCCGCGAGCGGCTGCAATCGC 1291
QY 628 TACTTCCATCGTACATGATCCCATCGGAAAGTCTCTCGAGGCAAGGTGTACCGA-AGCAG 687
DB 1292 TGCTTCCATCGTACATGATTCCTCTCGGATCATGTTGCTCGATCAGATGCCCTCACGT 1351
QY 688 CCAATGGTAAAGTTGACCGGAAGAACTCTCTCGAGGCAAGGTGTACCGA-AGCAG 746
DB 1352 CCAACGCAAGTTGGATCGCAAGAGCTCGCTCGACAAAGCCGGTCAATCCCAACAATTG 1411
QY 747 CAGACAGCAGCGCGGTTACCGACATTTCCCATCAGTGAGGTGCAAGTC-ATTCTTTGCGA 805
DB 1412 CCGCAAGCAGTTGGACTTTGTGGCGCGCACGCGGAAATCGAGTTCGGTCTCTGCGA 1471
QY 806 AGAAGCCACTGAGGTGTTGGCATGAAGTTGACATTTACCGATCACTTCTTCAATCTCGG 865
DB 1472 AGAATTACGATCTACTAGCGCTCAGGTCGCGCATTAACAGAACTTCTCAGATTGGG 1531
QY 866 TGGACACTCTCTTGGCCACGAAGTCAATTTCTCGATGACCAACGACTCAAGTTCG 925
DB 1532 CGGCCATTCGCTGCTGGCCACGAAGTGAAGCGACGCTCTAAGTCGAGAGTGGACGCCG 1591
QY 926 TATCACTGTCAAGGATGCTTTGACCATCTGTTATTTGGGATCTAGCATCTCTCATCCG 985
DB 1592 TGTCACTGTGAAGCAGATCTTTGACCAAGCAGTACTTGTGATCTTGTCTTCTTATTCG 1651
QY 986 TCAAGG 991
DB 1652 TCAAGG 1657

RESULT 10
AR050556
LOCUS AR050556 1713 bp DNA linear PAT 29-SEP-1999

DEFINITION Sequence 4 from patent US 5827706.
ACCESSION AR050556.
VERSION AR050556.1 GI:5973281
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1713)
AUTHORS Leitner, E., Schneider, E., Schoergendorfer, K. and Weber, G.
TITLE Cyclosporin synthetase
JOURNAL Patent: US 5827706-A 4 27-OCT-1998;
FEATURES
Location/Qualifiers
1. 1713
source /organism="unknown"
BASE COUNT 389 a 491 c 469 g 364 t
ORIGIN

Query Match 18.9%; Score 189.2; DB 6; Length 1713;
Best Local Similarity 59.5%; Pred. No. 2.3e-46;
Matches 396; Conservative 0; Mismatches 258; Indels 12; Gaps 4;
QY 338 GCTATCAACCATTCGGTCCAGCGCGGAGGCGACTCATCATATCCGTTCCCGACATCTT 397
DB 992 GATATCGCCACTCAATCACGGGGAAGGAATGCCCTGCTCTCTCAGTGGCGGACCTGAT 1051
QY 398 TCGCATTTGCTGGGAAGCGGGTTCCTGTCGAGGTTCAGTTCGACGACGAGTGGTCTCA 457
DB 1052 TGAGATTGGTAAGGGATCGGCTTCCAAAGTTGAGACCAGCTGGGCTCGACACACTCCCA 1111
QY 458 GAATGGTGCAATGGACGCTGTTTCCATCATTTGTTGCTCCCA-----GGCGGTAC 508
DB 1112 GCGCGGCGGACTCGATGCTGTTTCCACCGATTTCGAAACCAAGACACTCGGGTCAATG 1171
QY 509 TCTGGTCAACTTCTACGACCATCACTTCGAGGGTCTGATCTC-CTCACCAATCGAC 567
DB 1172 CATGTTTCAGGTTCCCAACTGAACACAGGGGCGGCTTTCGACGAGTCTCAAGATCGCC 1231
QY 568 CCCTTCAGCACTGCAAAACCGTCGATCGCCATCGAAGTCCGGAGAGGCTTCGGTCT 627
DB 1232 CGCTACACCTTGGTTCAGAGCGCGGCTGGAGGCAAGGTCCGCGAGCGGCTGCAATCGC 1291
QY 628 TACTTCCATCGTACATGATCCCATCGGAAAGTCTCTCGAGGCAAGGTGTACCGA-AGCAG 687
DB 1292 TGCTTCCATCGTACATGATTCCTCTCGGATCATGTTGCTCGATCAGATGCCCTCACGT 1351
QY 688 CCAATGGTAAAGTTGACCGGAAGAACTCTCTCGAGGCAAGGTGTACCGA-AGCAG 746
DB 1352 CCAACGCAAGTTGGATCGCAAGAGCTCGCTCGACAAAGCCGGTCAATCCCAACAATTG 1411
QY 747 CAGACAGCAGCGCGGTTACCGACATTTCCCATCAGTGAGGTGCAAGTC-ATTCTTTGCGA 805
DB 1412 CCGCAAGCAGTTGGACTTTGTGGCGCGCACGCGGAAATCGAGTTCGGTCTCTGCGA 1471
QY 806 AGAAGCCACTGAGGTGTTGGCATGAAGTTGACATTTACCGATCACTTCTTCAATCTCGG 865
DB 1472 AGAATTACGATCTACTAGCGCTCAGGTCGCGCATTAACAGAACTTCTCAGATTGGG 1531
QY 866 TGGACACTCTCTTGGCCACGAAGTCAATTTCTCGATGACCAACGACTCAAGTTCG 925
DB 1532 CGGCCATTCGCTGCTGGCCACGAAGTGAAGCGACGCTCTAAGTCGAGAGTGGACGCCG 1591
QY 926 TATCACTGTCAAGGATGCTTTGACCATCTGTTATTTGGGATCTAGCATCTCTCATCCG 985
DB 1592 TGTCACTGTGAAGCAGATCTTTGACCAAGCAGTACTTGTGATCTTGTCTTCTTATTCG 1651
QY 986 TCAAGG 991
DB 1652 TCAAGG 1657

RESULT 11
COPEPSYNH
LOCUS COPEPSYNH 3155 bp DNA linear PLN 03-FEB-1998

```
DEFINITION C. oligospermum gene encoding synthetase.
ACCESSION X96559
VERSION X96559.1 GI:1770179
KEYWORDS peptide synthetase; peptidase synthetase.
SOURCE Cyndrotrichum oligospermum.
ORGANISM Cyndrotrichum oligospermum
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Chaetosphaeriales; mitosporic Chaetosphaeriales;
Cyndrotrichum.
REFERENCE 1 (bases 1 to 3155)
AUTHORS Bernhard, F.
TITLE Identification of genes encoding for peptide synthetases from
gram-negative bacteria and filamentous fungi
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3155)
AUTHORS Bernhard, F.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1996) F. Bernhard, Freie Universitaet Berlin,
Institute of Crystallography, Takustr. 6, 14195 Berlin, FRG
FEATURES
    Location/Qualifiers
        1..3155
            /organism="Cyndrotrichum oligospermum"
            /strain="NRR1 18230"
            /db_xref="taxon:72418"
            /clone_lib="pCBI1"
            <1..>3155
            /function="amino acid adenylate forming module"
            /note="SD2 214-103"
            /codon_start=2
            /product="peptidase synthetase"
            /protein_id="CAA65395.1"
            /db_xref="GI:1770180"
            /db_xref="SptREMBL:P97961"
            /translation="DPOQLVPVGMGLVVTGDLARGVYTDPLVDGGRFVQIRIKGE
SVRAYRGDRARYRPGKQIEFFGRMDQQIKRGHRELAEEVHAHLHDAVRVAVV
LPSTGQPEMIGVVRADETVQDLSLRAGVANSANWEEQFEIQTEKEIRNLQNL
EEFSDVLGEVGVADSFYDGLGSLMATKLAARSRLNARSVSKVEYDFQVPLADLAI
VIQRGSPHNPILTAPYSGPVEQFAQGLWFLQINLGASWYLMPLAARVLRPLHTE
ALTATLAEORHEILRTTFEERQGVGVVREHLEKRLVDIVPADLDGGLQLOPKQ
BQTSFDLASEPGWRSLLRIANDHVLISVMHHIISDQWSIDILIRRELGLYSALR
GYDQSQSLPPIQYRDFSVQKQAEQVAERQLEWYTTTTLADDSPAELTDLPRPT
VLSNGVAVQALDGLSLEKURAFRCATQYTFSAVLLAAFRATHYRUTGAEDATIGTP
IANRRPELENIIDFFVNTQCMRTIVDRDFTFTLMOQVRSRTTAAAFANQDPERIV
SALLPGSDTNRNPLVLMVLSHQKDLGLEGVGEVPEPTAAGTFDEFLHFLQE
VRLGGTVLFDLFPDPTIRGMVTFQEVHLHVDOPHTPITSLPLANGLAGLNCMG
LLDIERSDYPRESSLDVDFEQVAACPDATAVIDSSRLTYTOLDROSDEYAGHLRR
HMAETLVGLAPRSCQTIYAFGLKANLAYLPLDVNVPAAIEALISAVEGHKLVL
LGSDDAPKRVQDVEMVRVDDTLGYDILDYDANSVAVRPSPTSLAYMFTSGSTGKPK
GVMEHRAVLRLVKNVSVISLPMTPRVHLSNLGFDISVQEVYFALLNGTGLVCIDY
FTTLDLSNALETVMQEQVRVAMLTPLLKLCGLADIPSVLGLDLVLFNVGDRFDSRDAL
AAQALIRGVVNAVYGPTENGI"
BASE COUNT 683 a 914 c 859 g 699 t
ORIGIN
Query Match 15.7%; Score 157.2; DB 8; Length 3155;
Best Local Similarity 61.9%; Pred. No. 1.6e-36;
Matches 249; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 590 TCGATATCCCATCGAAGTCGCGGAGAGCTTCGGTCTTACTTCCATCGTACATGATCCC 649
    || || || || || || || || || || || || || || || || || || || ||
Db 442 TCAGACTGAAAAGAGATCCGTAATCGGTTGCGAATCTGCTACCGTCGATGTCGCC 501
    || || || || || || || || || || || || || || || || || || || ||
QY 650 ATCGAACATCGTGTCTGGACAGATCGCTCTCAACGCCAATGTAAGTTGACCGGAA 709
    || || || || || || || || || || || || || || || || || || || ||
Db 502 GGCTCGGATACATACTAGACCATGCCCTTAAATGCAATGCAAGGTGACCGGAA 561
    || || || || || || || || || || || || || || || || || || || ||
QY 710 GGAATCTCTCGAGGGCAAGGTTGTACCGAAGCAGCAGCAGCGCGCGTATCCGAC 769
    || || || || || || || || || || || || || || || || || || || ||
Db 562 AGACCTTCTCGAAGGGCTCAGACCGTCTCGAAGCTCAGAAGTTGCCATCTGCACGCGT 621
    || || || || || || || || || || || || || || || || || || || ||
QY 770 ATTTCCCATCATGAGTTCGAGTCAATCTTTTCGAGAAAGCAACCTAGGCTGTTGGCAT 829
    || || || || || || || || || || || || || || || || || || || ||
```

| | BASE COUNT | 3016 a | 2429 c | 2550 g | 2518 t | |
|-----------------------|------------|---|--------|-----------------|------------|--|
| | | | | | | <p> IACRPHADLEAVLGMFVNTLALRTHPAGNKFTTDFDLKEVKQTALEAYEHQDYPPEELV DQUGVQREMSRDLFTTLLVQNKQKOTLKNNDLLLRWMDLQHPISKFDLSLXVTHFS AELCYQFESTALFEEDTVORMAGLFTTLLVENTSSPDMELDQIPPILEERSFIDS COPLREETGYRLKQTLHLAEQQAKPDQPAVIVEDEEITYKELNGAVANVELIDR GVAKEETVYALMGRRSPPEMLGITYAILKAGGAYIPIDPYPEERIRFLDKSDSDILLI QSDNLDIPAFDQFVPLHNSQDGLKRLSNPNVEVYSDSLAYMYITSGSGRPGKVQV EHOSANFNLNSQFVPLHNSQDGLKRLSNPNVEVYSDSLAYMYITSGSGRPGKVQV EPMILKYIIEQQITAMHFVPVSMHLHAFLEYLKNKGPVITKTNLRKRVSGEGUQAGHLV SRCELLPDVTLNSGPGTEATVFAAFDFLDEKDDIPYPIKPIHVRVILYNQOK MFLPAGCIGELYIAGAGVAYLNRPETLEERFLDDPFIYFGERMYKTQDLARLPDQVQ EFLGRDQDQIKRYIRTEPEIEBAALRETEGVREAAVTVESEGAELCAEYASGLGRN EVKQKLTLLPGVMIYPAHIITEMBQWPTPSGDKRKALPAPDCAADRETYAPRNLTIE MKLSOLWEEVLKSGVPYIHDNFRDGHSHLKATALYSRIAKFVGQVPLKDVFAHPTV EGLASVISEGTESPYEAIKPAGKRETYPVSSAKRYVYLOQEDGGTGYNMPAVLEIE GKLLDTRMDAVFKKLLKRHESLRTFTTDDGEPVQIRHDEVPFRFTSVLGGCTQOE AASAFIOPEDLSQAPFRAGIVYKSDGRHULLVDMHHIISDGVSVNTLIRFEGELYAN QELPULHIQKYDAYWQEAFAKGDYTKTQEAYWLKQLEGLPVLDPADYARPVRSF AGDQFPTLDQELTAGHLKARENGSTLYMVLAAAYTALLARYSGQEDITVGSPIAGR PHKDLSEILMGFVNTLHARREBGGTLYMVLAAAYTALLARYSGQEDITVGSPIAGR VTRDMSRNPVEDAMLVQNTQOCTFLADILQIOPAGVSHSHAKFDITLOASEGDEOIH LKFEYXSTALFOEATIKQWYHLTKLISHIVKNENTLNNDILITHEERNKLNFNSG QENHKVQITIDFEKALQATPNAPALPNSKENTLYTOELDWSNGLAWLRORNGKPD TPVGMWMPFSIMAAIAGVYKSGPQVSRPQHLAYIYTSKGGTQPKGVYVHBHGKIA OEALEDPEGEIFAENAEYTSKGPQVSRPQHLAYIYTSKGGTQPKGVYVHBHGKIA NTLQWHRNAYTINTEVLYLQISFSDGFLTSMFTPLLSAKTAVLPDENAKDILSTK HDLSHRRNAYTINTEVLYLQISFSDGFLTSMFTPLLSAKTAVLPDENAKDILSTK ELANEGVPTENSVAITAMRHMERQKVCIGRPDNTENVLLINGDLOQIOPAGVSELGIA GTCIARGYNLPELTAKTVQHPYQPEKRMVRTGDAARNWADGTIEYLRMDQOVKIR GHRVETKIESVRRISGVEAVVLEATAPEATQELCAIYVAEODNENTMLRAELNKO LPLYMPATPIQLDLSPLSGNKLRRALPEPAFQNVRTFPHATSEMEOKLADIWGEV LGAEKIDDSFELGDSITKALQVRLHRIKQMSVKDLSFHSHTIKELAAVIRSD TSAQAPVEGEVQWSPQVWFQSODIKKEHHFNOSVMLHRSTSIOEDAKRLVHVRDI HDLARMVFTQSBQGWQVYVNRVSHSDALYSLOIINLKGSEVNEVESILIKHVIRDI QONINLEKPLQAGLFRDGDGDFLTAHHFVLDGVSHRVLLEDLATGYRQAVNGED IDLPKTSSEKATEKLADYAESRQLMKQLSYWREAEYQAKALPDQMDAQATANERK RDTVSFSLNKETEALKHNSRQTLQDMLLASIALRDMTKTQASVKLSHSGHR EDVLDGIVSRVWGFYAIPEVLKDSLDLTKQDBEYLVHLVTKTTFDLRCVPGEDIS GVGVLKYPPEKTYGTFDTADISFNLYGFQENSRTEATIDENTISFSPGLGDIS TSNRQETLIDISAMAAEGRKVMNTYETSRFQHTKIAQLCENCHHYLLKLDHCLNKT ETBKTTISDPEHLETDALQEIADMLSFH" </p> |
| Query Match | 7.2% | Score 71.6; | DB 1; | Length 10513; | | |
| Best-Local Similarity | 52.8% | Pred. No. 3e-10; | | | | |
| Matches | 208; | Conservative | 0; | Mismatches 174; | Indels 12; | Gaps 2; |
| QY | 603 | GAAGTCGGCAGAGGCTTCGGTCTTACTTCCATCGTACATGATCCCATCGAACATCGTT | 662 | | | |
| DB | 5552 | GAAGTCGCAGACAGCTTGAGACTTTGCTTCGGGCTACATGATCCTGCTCATATCATTT | 5611 | | | |
| QY | 663 | GTTCCTGGACAAGATGCGTCTCAACGCCAATGGTAAAGTTGACCGGAGGAACCTCTCTCGC | 722 | | | |
| DB | 5612 | GAGATGGAACAATGGCCGGTTACCCCGAGCGGAAAGCTCGACCGAAAGCGCTGCCT--- | 5668 | | | |
| QY | 723 | AGGGCAAGGTTGTACCGGACGACGACAGACAGCGCCCGTTACCGACATTTCCCATCAGT | 782 | | | |
| DB | 5669 | -----GCTCCTCAGCAGCGGCGAGATCGGGAAACCTACACAGCGCCGAGAAATCTGA | 5720 | | | |
| QY | 783 | GAGGTCCGAAGTCAATCTTTTCGGAAGAAGCCACATGAGGTGTTTGGCATGAAGGTGTGACAT | 842 | | | |
| DB | 5721 | CAGAGATGAAGCTTTCCAGCTGTG--GGAGGAAGTCTGNAAGCGGCCCTGTGCGCATTT | 5779 | | | |
| QY | 843 | ACCGATCACTTTCTTAATCTCGGTGGACACTCTCTCTTGGGCCACGAGGCTCATTTCTCGT | 902 | | | |
| DB | 5780 | CACGACAACCTTTTTCACCGGGCGCGCATTTCTCTAAAAGCCACCGCGCTTGTCTCCCG | 5839 | | | |
| QY | 903 | ATCGACCAACGACTCAAGGTCCGTATCACTGTCAAGSATGTCTTTGACCATCTCTGATTT | 962 | | | |
| DB | 5840 | ATCGCAAGGAATTCGGGTGTCGAAGTGCCGCTGAAGGATGTGTTCCGCCATCCGACGGTA | 5899 | | | |
| QY | 963 | GCGGATCTAGCATCTCTCATCCGTCAAGGCGTGG | 996 | | | |


```
Db 3530 -----GCTCCTGACGGAGCGGAGATCGGAAACCTACAGCACCAGAAATCTGA 3581
Qy 783 GAGGTGGAAGTCATTCTTTGCGAAGACCCACTGAGGTGTTTGGCATGAAGTTGACATT 842
Db 3582 CAGAGATGAAGCTTTCCAGCTGTG-GGAGGAAGTGTGAAAGACGGCCCTGTGCGGCATT 3640
Qy 843 ACCGATCACTTCTCAATCTCGGTGGACACTCTCTCTTGGCCACGAAGCTCATTTCTCGT 902
Db 3641 CACGACACATCTTTTGACCGGGGGCGCAATCTCTTAAGACCACCGCGCTGTGCTCCCGG 3700
Qy 903 ATCGACCAACAGCAACAAGGCCGATCACTGTCACGATGTCCTTTGACCATCTCTGTAATT 962
Db 3701 ATCGCCAAAGGAATTCGGCTGCTCAAGTGCCTGGAAGATGTGTTGCGCCCATCCGACGTA 3760
Qy 963 CGCGATCTAGATCTGTCATCCGTCGTCAGGCGCTGG 996
Db 3761 GAAGTTTGGCTGTCTATCATCAGCGAAGGAACGG 3794

RESULT 15
BSPEPSYN
LOCUS B.subtilis genes for peptide synthetase and penicillin binding
DEFINITION B.subtilis genes for peptide synthetase and penicillin binding
protein.
ACCESSION Z34883
VERSION 1
KEYWORDS penicillin binding protein; peptide-synthetase; pps operon.
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
Tognoni,A., Franchi,E., Magistrelli,C., Colombo,E., Cosmina,P. and
Grandi,G.
TITLE A putative new peptide synthase operon in Bacillus subtilis:
partial characterization
JOURNAL Microbiology 141 (Pt 3), 645-648 (1995)
MEDLINE 95227362
PUBMED 7711903
REFERENCE 2 (bases 1 to 39822)
AUTHORS Grandi,G.
DIRECT SUBMISSION
Submitted (27-JUN-1994) Grandi G., Enricerche S.P.A., Genetic
Engineering and Microbiology, Via F. Maritano,26, S. Donato
Milanese (MI), ITALY, 20097
REVISED BY [3]
3 (bases 1 to 39822)
de Ferra,F. and Tognoni,A.
DIRECT SUBMISSION
Submitted (24-JAN-1997) F. de Ferra, Enricerche S.P.A.,
Environmental Biotechnology Laboratories, Via F. Maritano,26, S.
Donato Milanese (MI), ITALY, 20097
REVISED BY author 27-JAN-1997
On Jan 29, 1997 this sequence version replaced gi:509465.
FEATURES
source
1..39822
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
<1..129
/codon_start=1
/transl_table=11
/label=CDS1
/product="hypothetical protein"
/protein_id="CA84365.1"
/db_xref="GI:509466"
/db_xref="SWISS-PROT:P39840"
/translation="PTVTNVAVDLPSSLTGLQVLEPGEETAKSSITIELNHQ"
147..156
163..1638
/gene="ppb"
163..1638
/gene="ppb"
/codon_start=1
CDS
RBS
gene
CDS
```

```
/transl_table=11
/label=CDS2
/product="putative penicillin binding protein"
/protein_id="CA84366.1"
/db_xref="GI:509467"
/db_xref="SWISS-PROT:P39844"
/translation="MKKSIKLYVAVLLLVVVASVYPVHQAALAAAEKODALSGQIDKIL
ADHPALGAMAGITVRSAGTGAVALYEHSGDTRMRPASLKLTTAAALSLVGENYSFT
TEVTDGTLGKGLNGLYLKGKGPDLTPSPDFDKMAEILKHSVKVYIKGNLIGDWTW
HDMRLSPDMPWSDEYTYGAPISALTASPNEYDAGTVIYEVTPNOKGEPRAVSVS
PKTDYITIKNDAKTTAAGSEKDLTIERHGTNTITIEGSPVDANKTKEWLSVNEPAG
YALDLFKOSLKKOGITVKGDKIKTGEAPSSDVLSSHRSMPLSKLFVFPFMKUSNNGHAE
VLVKEMKVRKGGSEKLEVLNSTLPEFGSDSKSLVLRDGSIGSHIDAQSDQSLQ
LLYDIQDSWFSAYLNSLPVAGNPRMVGTTLRMRKGTPAQGRKARTKSGLSLSTVSSL
SGYAEKSGKKLVFSILLNGLIIDEEDGKDIEDQIAVILANQ"
1648..1678
2009..2013
2021..9706
/gene="ppsl"
2021..9706
/codon_start=1
/transl_table=11
/label=CDS3
/product="putative fengycin synthetase ORF1"
/protein_id="CA84360.1"
/db_xref="GI:580910"
/db_xref="SWISS-PROT:P39845"
/translation="MSBHTYSLTHAQRVWVTELEPDTSCINLTACVKPKGNIELDT
LEGALNHSISRDAIRPOLLEGELEPLRLTEYKYVPLRIDFSNVEIEIOMIOD
QASIPKLFNSPLFOFYLRLIDSDHEVWLFAKFHILINDGISLVNMGNOIDLKQMKK
KDLPOPEPSYLSYTEKSOYQSPFAKDRLPWTQTFEPLHYLSLADTISLQSKS
TSASRTIILSPDLQITRIFCEEHKINISLFMASFYICISRTISKKDLAIGTYYN
RGEAKEMIGMEVSSLPITRTVDPTDFLSFVRTIGREQLSVMRHORFPYNLLVNL
RNEQKOLNHLIGISMOYQPLQHNADDDEYETALFYSGYTANELSYQIERIDNGTQ
LNFDYQNTLFSLIEDIKRIQSHLLTILENALKHPHSFIRELDMTNRKOKLIFYENKT
EAVSPKATLHGLFERQAATPERLAIRFSGGSIYAEIDMVASRLAAHAAAGVTNE
SIVGLSERSPDLIAVLAVLKAGGATLPDPAYPKRLSYMLKDSGLLITOPGCS
APNFSGETLEVDMTSLASAEKNEHFTPDAGGSYAYIYTSGGTQGPQGAVERHQAV
SFLTGMQHOFPILSEDDIVMYKTSFSDASVQMLFWNSLGSASAYLLPQMKDSALV
QALHOENVTTAFHETIPAMNLSFLDQAEIERISDRTSLKRVFAGGEPAPRTAAARSVL
POVSLHGYGTEATDAAYVLDPRDRDLRIPIGKVPVGARLYVLDPLHAPVQPSG
VAGELYTAGAVAGYLNRPALTTEPFLEDPEYGERMYKTGDVARNLPDGNVEFLGR
TDDQVATRGVRIEPEGEALARSLEGVREAAVTVDSDGEPELCAYVEGLQORNEVRAQ
LQRLLPGYMVPAYIMEQWPVTPSGKLDNRNALPAPGAADAETITAPRVNTEKLSQ
LWEDVLKGPVGIHDMNPFDRGHSKLKATALVRIAKFEFQVPLKDFAHPTVEGLAT
VIREGTDSPYEAIKPAEKQETYPVSSAQKRIYVLOQLEDGGTGYNMPAVLEGLKAP
ERMERAFKELIKRHESLTSFEQDAGDPVORIHDEVPFTLQTTVLGERTEQEAAAF
IKPFDLSOAPLFRQIVKISDERHLLVDMHHIISDGSVSNILIRFEGLYNNRNLPA
LRIOXDYAVMREGFKTGDAYKTOEAYWLKQLEGELPVLDPADHARPPVRSFAGDKV
SFTLDQEVASGLHKLARENGSTLYMVLAAATAPLSRLSGOEDIIVGSPJAGRPHKDL
EPILGMFVNTLALTRPEGGKPFVQLQEVRETALEAFEHODYPFEELVDKLELTRDM
SRNPVDFAMF ILQNVKQIDILREIKVRPANFAHHSILFDITLIATISGICEMEF
STEVFLKATIERWADHFIEPLHEALSTPETAHILNLSDEKQKIVFEFNKQVFEFA
OKDI PFHRIFEAKAENPEHIAVIDNETEISYRLNERANRLARTLONRKGKPEVTAV
LAKRSIDAIVGLAVMKAGGVIPIDAHYPAKRIEYILRSGADILLQRELKHLISN
SPEKSMHIFLDDEGSEESNLSNPAPPEEYIIYTSGTGAPGVILVYONFTH
AALARQIYELDRKRVLLQIASFSDVFDGLARTLNGTTLIVPDETRLEPAEYI
QSMRIINSYVMESTPALIPVMEYVYNQKLPDLDLILGSDMVAQDRKTLTDREFG
VAGELCTIGGAGVAKGYHKHPDLTQMKTENPFVSGRBYTGDRAWLPNGTTRLLGR
MXYQVKINGYRIETEEIESVLLQTLGLVRENAVQHDKKQAGLAAYIVPSDVNTNAL
RAALTKEPAMITPAYILIPNLPLTKNGLDNRNALPAPNNVLSRPTAPVNDLQKTM
AYIWEDVLSMRVGIHDSFELGDSIKALQVAARAAAGESMTIRDLFRYSITQELC
GHITPLASQDQGAEGEALPTIQRRFFQGHAFHYHNSQWLFSEKGFNANLHL
ALKRIETEHDAIRMIFQRQNGHVIQFNRIHNDHELFLGLYISDMTWSRLLEDLAA
KLAAEETVTSQKMNVEKPLQOAGLFTAGEGDHLIALHLLVIDGYSWRTLLLEDLAA
YQALSEKKEIQLPPKTDLSYADGLTQAESKQLSEKTYWQTLIDTAFTAPLPKDI
NVPDKLQMSDAAAVLISDWTGKFTSQATGTADANELLTALGMALESEWGHQOI
VISTEGREGHVFNIDISRTVEGFTSIYPLDMDGIPPEPFDQARTRKTKMLLR
VPNKGTGYLLTHIGELRHKEPEVSFNLQGFSEKEVFTQLSYQPRYIAGERER
EYELDINALITGLRHVKAVYTOVFSKHSTECFMDRPHRLIETIEHCSQKAREKTL
SDFSNNKLTLSLSSIEDLVKDL"
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:28:42 ; Search time 271.5 Seconds
(without alignments)
8302.948 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000

Perfect score: 1001

Sequence: 1 aacctgttttctacctcg.....tcgtcaaggctgggtttg 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 1001 | 100.0 | 11212 | 21 | AAA58762 |
| 2 | 282.6 | 28.2 | 9633 | 22 | AAF79702 |
| 3 | 240.2 | 24.0 | 46899 | 15 | AAQ54386 |
| 4 | 187.6 | 18.7 | 1713 | 15 | AAQ54389 |
| 5 | 53 | 5.3 | 13029 | 23 | AA551470 |
| 6 | 50 | 5.0 | 41599 | 22 | AAI66165 |
| 7 | 49.6 | 5.0 | 1200 | 20 | AAV69559 |
| 8 | 49.2 | 4.9 | 7178 | 21 | AAA59145 |
| 9 | 48.2 | 4.8 | 37856 | 21 | AAA11992 |

| | | | | | | |
|------|------|-----|---------|----|----------|--------------------|
| 10 | 48 | 4.8 | 18660 | 21 | AAA58472 | Nucleotide sequenc |
| 11 | 46.8 | 4.7 | 6465 | 24 | ABK74875 | Bacillus lichenifo |
| 12 | 46.8 | 4.7 | 8268 | 24 | ABK74878 | Bacillus lichenifo |
| 13 | 45.4 | 4.5 | 4615 | 22 | AAF90034 | Nucleotide sequenc |
| 14 | 45.4 | 4.5 | 34071 | 22 | AAF90033 | Nucleotide sequenc |
| c 15 | 45.4 | 4.5 | 42717 | 22 | AAF90032 | Nucleotide sequenc |
| 16 | 42.8 | 4.3 | 88421 | 24 | AAI40781 | 88421nt genomic DN |
| c 17 | 42 | 4.2 | 2365589 | 24 | ABA90521 | Genomic sequenc o |
| 18 | 41.8 | 4.2 | 58857 | 21 | AAA58471 | Nucleotide sequenc |
| 19 | 40.4 | 4.0 | 7347 | 23 | AA554136 | Pseudomonas aerugi |
| 20 | 39.8 | 4.0 | 3798 | 24 | ABK74884 | Bacillus lichenifo |
| 21 | 39.8 | 4.0 | 3798 | 24 | ABK74902 | Bacillus lichenifo |
| 22 | 39.8 | 4.0 | 5451 | 22 | AAF26319 | Pseudomonas sp lip |
| 23 | 39.4 | 3.9 | 68750 | 21 | AAZ55887 | Sorangium cellulos |
| 24 | 39.4 | 3.9 | 71989 | 21 | AAA29349 | Sorangium cellulos |
| c 25 | 39.2 | 3.9 | 1172 | 20 | AAV69527 | Soil bacteria pept |
| c 26 | 39.2 | 3.9 | 4620 | 22 | AAF26318 | Pseudomonas sp lip |
| c 27 | 38.8 | 3.9 | 1178 | 20 | AAV69562 | Soil derived pepti |
| 28 | 38.4 | 3.8 | 31122 | 14 | AAQ40706 | Bacillus subtilis |
| 29 | 37.8 | 3.8 | 3849 | 22 | AAF25795 | S. chrysomallus ac |
| c 30 | 37.6 | 3.8 | 4403765 | 22 | AAI99683 | Mycobacterium tube |
| c 31 | 37.6 | 3.8 | 4411529 | 22 | AAI99682 | Mycobacterium tube |
| 32 | 37.4 | 3.7 | 759 | 22 | AAH52819 | S. epidermidis ope |
| 33 | 37.4 | 3.7 | 2301 | 22 | AAH52513 | S. epidermidis ope |
| c 34 | 37.4 | 3.7 | 2634 | 22 | AAH54825 | S. epidermidis gen |
| c 35 | 37.4 | 3.7 | 2744 | 22 | AAH54705 | S. epidermidis gen |
| c 36 | 37.4 | 3.7 | 3321 | 22 | AAH54694 | S. epidermidis gen |
| c 37 | 37.4 | 3.7 | 3602 | 22 | AAH54066 | S. epidermidis gen |
| c 38 | 37.4 | 3.7 | 3825 | 22 | AAH54435 | S. epidermidis gen |
| 39 | 37.4 | 3.7 | 6968 | 22 | AAH52883 | S. epidermidis ope |
| 40 | 36.8 | 3.7 | 1716 | 22 | AAF26315 | Pseudomonas sp lip |
| 41 | 36.8 | 3.7 | 7110 | 22 | AAF81361 | Quorum sensing con |
| 42 | 36.4 | 3.6 | 11601 | 12 | AAQ13608 | ACV synthetase gen |
| 43 | 36.4 | 3.6 | 13058 | 14 | AAQ48231 | Vector containing |
| c 44 | 36.2 | 3.6 | 4541 | 20 | AAV72100 | Human RON receptor |
| c 45 | 36.2 | 3.6 | 4541 | 24 | ABN97290 | Gene #3788 used to |

ALIGNMENTS

RESULT 1
AAA58762
ID AAA58762 standard; DNA; 11212 BP.
XX
AC AAA58762;
XX
DT 20-OCT-2000 (first entry)
XX
DE DNA encoding a cyclohexadepsipeptide synthetase.
XX
KW Cyclohexadepsipeptide synthetase; filamentous fungal cell;
KW Cyclohexadepsipeptide; antibiotic; ss.
XX
OS Fusarium venenatum.
XX
FH Key Location/Qualifiers
FT CDS 270..9659
FT /*tag= a
FT /product= "cyclohexadepsipeptide synthetase"
XX
PN WO200042203-A2.
PD 20-JUL-2000.
XX
PF 13-JAN-2000; 2000WO-US00913.
XX
PR 13-JAN-1999; 99US-0229862.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX
PI Berka RM, Rey MW, Yoder WT;
XX

DR WPI: 2000-482833/42.
 XX P-PSDB; AAB07427.
 PT Producing a heterologous polypeptide for production of antibiotics
 PT comprises cultivating a mutant of a parent filamentous fungal cell
 PT comprising a nucleic acid sequence encoding cyclohexadepsipeptide -
 XX
 PS Claim 55; Fig 1A-I; 76pp; English.
 XX
 CC The present sequence encodes a cyclohexadepsipeptide synthetase
 CC polypeptide. The specification describes a method for producing a
 CC heterologous polypeptide. The method comprises cultivating a
 CC mutant of a parent filamentous fungal cell, which produces less
 CC cyclohexadepsipeptide than the parent filamentous fungal cell when
 CC cultured under the same conditions. The method if used for the
 CC production of biologically active compounds e.g. antibiotics.
 XX
 SQ Sequence 11212 BP; 2808 A; 2844 C; 2658 G; 2901 T; 1 other;
 Query Match 100.0%; Score 1001; DB 21; Length 11212;
 Best Local Similarity 100.0%; Pred. No. 4.9e-313;
 Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AACCTGCTTTCTTCCACCTCGTTGAAGACAGAGTTTCCAGGCTGCTGGTGGACATCTTGAGA 60
 DB 7000 AACCTGCTTTCTTCCACCTCGTTGAAGACAGAGTTTCCAGGCTGCTGGTGGACATCTTGAGA 7059
 QY 61 TCCTGCCAAAGAACATGGAAGCTGTGAATGAGCTCAGTGGGTATCGATATGCCGCTGTTG 120
 DB 7060 TCCTGCCAAAGAACATGGAAGCTGTGAATGAGCTCAGTGGGTATCGATATGCCGCTGTTG 7119
 QY 121 TGCAGCTTCGGGGTCTACTTGGAGATGAGCTGTGCTTCCGGTTGAAAGATGACTGGA 180
 DB 7120 TGCAGCTTCGGGGTCTACTTGGAGATGAGCTGTGCTTCCGGTTGAAAGATGACTGGA 7179
 QY 181 TCGACTTTCAAGGAATCAATTGAACAGAGACTGCTGGTGGCTTCTCAAGTCTTCAG 240
 DB 7180 TCGACTTTCAAGGAATCAATTGAACAGAGACTGCTGGTGGCTTCTCAAGTCTTCAG 7239
 QY 241 ATGCTGCTATCGGAGTCAAGCAAAATTCCTTTCGAATACAGGCTTTCGAAGACAGG 300
 DB 7240 ATGCTGCTATCGGAGTCAAGCAAAATTCCTTTCGAATACAGGCTTTCGAAGACAGG 7299
 QY 301 TCGTCTGCTTCCCTCAATAGCAATCATGAGTGGGAGCTATCAACCATTCGCTCCAGG 360
 DB 7300 TCGTCTGCTTCCCTCAATAGCAATCATGAGTGGGAGCTATCAACCATTCGCTCCAGG 7359
 QY 361 CCGAGGCGGACTCATCACTATCCGTTCCCGACATCTTTCGATTCGCTGGGGAAGCCGGT 420
 DB 7360 CCGAGGCGGACTCATCACTATCCGTTCCCGACATCTTTCGATTCGCTGGGGAAGCCGGT 7419
 QY 421 TCGGTGTCGAGGTCAGTTTCGACGACAGTGGTCTCAGATGTTGCAATTCGACGCTGTTT 480
 DB 7420 TCGGTGTCGAGGTCAGTTTCGACGACAGTGGTCTCAGATGTTGCAATTCGACGCTGTTT 7479
 QY 481 TCCATCATCTTCTGCTCCAGGCGGTACTCTGCTCACTTCTTCCAGGACCATCACCTTC 540
 DB 7480 TCCATCATCTTCTGCTCCAGGCGGTACTCTGCTCACTTCTTCCAGGACCATCACCTTC 7539
 QY 541 GAGGGTCTGATCTCTTCAACCAATCGACCCCTTCAGGAGCTGCAAAACCGTCGATCGCCA 600
 DB 7540 GAGGGTCTGATCTCTTCAACCAATCGACCCCTTCAGGAGCTGCAAAACCGTCGATCGCCA 7599
 QY 601 TCGAAGTCCCGAGAGGCTTCGCTTACTTCCATCGTACATGATCCCATCGAATCG 660
 DB 7600 TCGAAGTCCCGAGAGGCTTCGCTTACTTCCATCGTACATGATCCCATCGAATCG 7659
 QY 661 TTGTTCTGGACAGATGCTCTCAACGCCAATGTAAGTTGACCGGAAGACTCTCTC 720
 DB 7660 TTGTTCTGGACAGATGCTCTCAACGCCAATGTAAGTTGACCGGAAGACTCTCTC 7719
 QY 721 GCAGGGCAAGAGTTGTACCGAAGCAGCAGACAGCGCGCTTACCGACATTTCCCATCA 780

DB 7720 GCAGGGCAAGAGTTGTATCCCGAAGCAGACAGCAGCGCCGTTACCGACATTTCCCATCA 7779
 QY 781 GTGAGGTGCGAAGTCAATTTCTTGGGAAGCAAGCACTGAGGTGTTGGCATGAAGTTGACA 840
 DB 7780 GTGAGGTGCGAAGTCAATTTCTTGGCAAGCAAGCACTGAGGTGTTGGCATGAAGTTGACA 7839
 QY 841 TTACCGATCACTTCTTCAATCTCGGTGGACACTCTCTTTCGCCACGAAAGCTCATTTCTC 900
 DB 7840 TTACCGATCACTTCTTCAATCTCGGTGGACACTCTCTTTCGCCACGAAAGCTCATTTCTC 7899
 QY 901 GTATCGACCAACGACTCAAGTCCGTATCAAGTATCAAGTATGTTGACCATCTGTAT 960
 DB 7900 GTATCGACCAACGACTCAAGTCCGTATCAAGTATGTTGACCATCTGTAT 7959
 QY 961 TTGGGATCTAGCATCTGTATCGTCAAGGCTGGGTTG 1001
 DB 7960 TTGGGATCTAGCATCTGTATCGTCAAGGCTGGGTTG 8000
 RESULT 2
 AAF79702
 ID AAF79702 standard; DNA; 9633 BP.
 XX AAF79702;
 AC AAF79702;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 DE Mycellia sterilia cyclic depsipeptide synthase DNA.
 XX
 KW Mycellia sterilia; cyclic depsipeptide synthase; anthelmintic;
 KW PF1022; ds.
 XX
 OS Mycellia sterilia.
 OS
 XX WO200118179-A1.
 PN
 XX 15-MAR-2001.
 PD
 XX 07-SEP-2000; 2000WO-JP06103.
 PF
 XX 07-SEP-1999; 99JP-0253040.
 PR
 XX 06-APR-2000; 2000JP-0104291.
 XX
 PA (MEIJ) MEIJI SEIKA KAISHA LTD.
 XX
 PI Midoh N, Okakura K, Miyamoto K, Watanabe M, Yanai K, Yasutake T;
 PI Aihara S, Futamura T, Kleinkauf H, Murakami T;
 PI
 DR WPI: 2001-265970/27.
 DR P-PSDB; AAB73959.
 XX
 Novel cyclic depsipeptide synthase and gene encoding it for efficient
 production of anthelmintic substance Pf1022
 PT
 PT
 XX
 PS Claim 3; Page 30-58; 92pp; Japanese.
 XX
 CC The present sequence encodes a cyclic depsipeptide synthase from Mycellia
 CC sterilia. The protein encoded by this sequence, or a protein containing
 CC an addition, deletion and/or substitution of one or more amino acid
 CC residues is useful for the efficient production of the anthelmintic
 CC Pf1022 (cyclo(D-lactyl-L-N-methylleucyl-D-3-phenyllactyl-L-N-
 CC methylleucyl-D-lactyl-L-N-methylleucyl-D-3-phenyllactyl-L-N-
 CC methylleucyl)).
 XX
 SQ Sequence 9633 BP; 2318 A; 2834 C; 2462 G; 2019 T; 0 other;
 Query Match 28.2%; Score 282.6; DB 22; Length 9633;
 Best Local Similarity 59.5%; Pred. No. 3.2e-80;
 Matches 618; Conservative 0; Mismatches 364; Indels 57; Gaps 6;
 QY 20 GTTGAAGACAGGTTTCCAGGCTGTTGGTGAACATGTTGAGATCTGCTGCAAGAACATGGA 79
 DB 6900 GCTGCAGGCCAGCTTCCCGATCGATCAAGCAGCTGAGATCTCTCCGGAACATGCG 6959

```
QY      80  AGCTGTGAATGAGCTCAGTGGTATCGATATCGCCTGTGTGTGACGTTTCGGGTTCACT 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      6960 CGCCACGAACGAGCTGAGCGGTACCGGTATACAGCGCTATTCACGTACGGGCCGAGA 7019

QY      140  TGGAGA---TGAGTTGTGCTTCCGGTTGAGAAAGATGACTGGATCGACTTTCAAGCGAA 196
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7020 GGAACAGTCGGGGCCCGTGTATCCGATCCAAAGTGAACGACTGGATCGACTTTCAAGCCTC 7079

QY      197  TCAATTGAACCAAGATCACTGGGTGACCTTCTCAAG---TCTTCAGATGTCGTATATC 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7080 AGGCATTCAGCCCGCCGCTTCTCCGACTCTACAGCGCTCGGCAGACGGCGCACCGT 7139

QY      254  GGCAGTCAGCAAAATTCCTTCGAAATCACGCCCTTTGAAAGACAGGTCGTTCGCTTCCT 313
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7140 CGCCGTCAGCAACATCCCTACAGCAAGACGATTAGAACGCCATGTCGTGAGTCCCT 7199

QY      314  CAATAGCAACATCGATGAGTGGCAGC-----TATCAAC 346
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7200 TGACAATAACACAGGAGGATACGCATAGACACAGACGCGCGGCTTGGATCTCGC 7259

QY      347  CATTCGGTCCAGCGCGGAGGAGCTCATCTATCCGTTCGCCGACATCTTTCGCATGCG 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7260 CGTCCGCTCCAGGCGGAGCGCTGCACGTCCCTCTCCGTGACCGATCTTGTGCACCTCG 7319

QY      407  TGGGGAAGCGGGTTCCGTGTCGAGGTGAGTTCTGCAGGACAGTGTCTCAGAATGGTGC 466
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7320 GGAAGACGCGGCTTTCGCGTAGAAGTCAGCGCAGCGCGGAGTGTCTCAAAGCGGCG 7379

QY      467  ATTGACGCTGTTTTCATCATTTGTGCT-----CCCAAGGCGTACTCTGCT 514
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7380 GCTCGATCGCGTCTTTCACCGCTATAATTTGCCACTCAAAACAATAGTTCGCGTCTGAT 7439

QY      515  CAACCTTCTACGGACCAT---CACCTTCGAGGGTCTGATCTCCTCACCAATCGACCCCT 571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7440 TCAGTTCCTACGAAGATGGCCAGCGGAAGATCCGCCACTCTGACAAACCGACCAT 7499

QY      572  TCAGCGACTGCAAAACCTGCTATCGCCATCGAAGTCCGCGAGAGGCTTCGGTCCCTTACT 631
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7500 ACAGCGTCTGCAGAGCGCGGTTCGCATCACAGATCCGGAACAGCTGAAGGCGGTGCT 7559

QY      632  TCCATCGTACATGATCCCATCAACATCGTTGTTCTGGACAGATGCCCTCTCAAGCCCA 691
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7560 CCGGTACATACATGATCCCGCTCCCGCATCGTGGTCTATAGACACAGATGCCCTCAATGCCAA 7619

QY      692  TGTAAAGTTGACCGGAGGAACTCTCTCGCAGGGCAAGGTTGTACCGAAGCAGCAGAC 751
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7620 TGGCAAGTTCGACCGGAAGAACTTACCAAGAGGGCCCAATCGCGCCGAATCTCAGGC 7679

QY      752  AGCAGCGCGGTTTACC-----GACATTTCCCATCAGTGAAGTCGAGTCAATCTTTG 802
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7680 GGTCCCGCAAAACCCGTCAAACAGTCGATCCGTTCTGTCACCTTGGAAAGCCATTTATG 7739

QY      803  CGAAGAAGCACTGAGGTGTTGGATGAGAGTTGACATTACCGATCACTTCTCAATCT 862
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7740 TGAGGAGTTTCGGGAGGTCGTGGGATGGAAGTCGGCGTGAACGACCACTTCTTCAACT 7799

QY      863  CGGTGACACTCTCTCTTGGCCACGAAGCTCAATTTCTGATCGACCAACGACTCAAGT 922
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7800 AGCGGACACTCTCTCTTGGCCCAAGAACTCTCGGGCGTCTCAGTCTCGCTTAACGG 7859

QY      923  CGGTATCACTGTCAAGAGTGTCTTTGACCATCTCTGATTTGCGGATCTAGCATCTGTCT 982
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7860 TCGTGTGTGTGTGAGGATGTGTTCGACCAAGCTGTGATTTCCGACCTCGCAGTCACT 7919

QY      983  CCGTCAAGGGCTGGGTTTG 1001
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      7920 CCGCCAGGAGCTGACCTTG 7938
```

RESULT 3

AAQ54386

ID AAQ54386 standard; DNA; 46899 BP.

```
XX      AAQ54386;
AC      08-JUL-1994 (first entry)
DT
DE      T. niveum Cyclosporin synthetase gene.
KW      Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
      T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase; ss.
KW      Tolypocladium niveum.
OS
XX
FH      Key Location/Qualifiers
      CDS 885..46730
      FT /*tag= a
      FT /product= Cyclosporin synthetase
      FT misc_feature 40239..43129
      FT /*tag= b
      FT misc_feature 37781..40244
      FT /*tag= d
      FT /note= "Sali restriction fragment, preferred
      FT fragment, Claim 4"
      FT /note= "Sali restriction fragment, preferred
      FT fragment, Claim 5"
XX
PN      EP578616-A.
XX
PD      12-JAN-1994.
XX
PF      05-JUL-1993; 93EP-0810474.
XX
PR      09-JUL-1992; 92AT-0001403.
      08-MAR-1993; 93AT-0000437.
      29-APR-1993; 93CH-0001310.
      04-MAY-1993; 93CH-0001375.
XX
PA      (SANO ) SANDOZ LTD.
      (SANO ) SANDOZ PATENT GMBH.
      (SANO ) SANDOZ-ERFINDUNGEN VERW GES MBH.
PI      Leitner E, Schneider E, Schoergendorfer K, Weber G;
XX
DR      WPI; 1994-010432/02.
      P-PSDB; AAR4929.
XX
PT      Isolated DNA sequence - which codes for enzyme having cyclosporin
      synthetase like activity
XX
PS      Claim 6; Page 17-41; 93pp; English.
XX
CC      This sequence encodes an enzyme which has cyclosporin synthetase-
      like activity. This sequence was isolated from Tolypocladium niveum
      (formerly known as T. inflatum GAMS). The enzyme encoded by this
      sequence catalyses the peptide biosynthesis of cyclosporins and
      structurally related molecules. This sequence may be used for the
      production of cyclosporin by transforming a vector containing this
      sequence in to a recombinant host. This allows effective production
      of antibiotic cyclosporin or its derivatives.
XX
SQ      Sequence 46899 BP; 10651 A; 13513 C; 12509 G; 10226 T; 0 other;
      Query Match 24.0%; Score 240.2; DB 15; Length 46899;
      Best Local Similarity 55.0%; Pred. No. 4.2e-66;
      Matches 563; Conservative 0; Mismatches 418; Indels 42; Gaps 3;
```

```
QY      1  AACCTGCTTTCTTACCTCGTTGAAGACAGAGGTTTCCAGGTCGTGGTGAACATGTTGAGA 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      21076 ACCCTGCAPTTTTATACATCTTTCGTACGCTTGGTGGAGAAGATCAAGCAGCTGAAA 21135
```

```
QY      61  TCCTGCCAAGAACATGGAAGCTGTGAATGAGCTCAGTCCGTATCGATATCGCGCTGTTG 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      21136 TTCTCCCAAGACCATGAAGGCTTACCACGAGCTCAGCAAGTACCGATATGCGCGAGTAC 21195
```

| | | | |
|----|-------|--|-------|
| Qy | 121 | TCGACGTTCCGGGTTTCACATTGGAGATGAGCTTGCTCCGGTTGAGAAAGATGACTGGA | 180 |
| Db | 21196 | TAGATTGGCTGGCTCGAGAGAACAAATCAACTATACACAAGTCTCTCCCAACGCCTGGA | 21255 |
| Qy | 181 | TCGACATTTCAAGGGAATCAATTGAACCAAGACTCACTGGGTCACTTCTCAAGTCTTCAG | 240 |
| Db | 21256 | TAGACTTTGGGCGAGACGGTCTCGACCGGACAGCCCTCATCAACTTGCTGAAGGAGCACA | 21315 |
| Qy | 241 | ATGCTGCTATCATGGCAGTCAAGCAAAATTCCTTTGCAAAATCAGGCGCTTGAAAGACAGG | 300 |
| Db | 21316 | AGGATCCGGGACCGTCGCTATCCGTTAATATCCCTACAGCAAGACCAATTGTTGAGCGGT | 21375 |
| Qy | 301 | TCGTCGTTCCCTCAATAGCAACATCGATGAGTGCGCA | 337 |
| Db | 21376 | TTGTCAAACAGTCACTGAGCGAGGATGATGAGGAAGGCGAGAACTCACTGACCGGAT | 21435 |
| Qy | 338 | -----GCTATCAACATTCGGTCCAGCGCGAGGCGGACTCATCACTATCCGTTCCCG | 390 |
| Db | 21436 | CAGCTTGGGTTGACGGCTCCGGATGGCGGCTCAAAAGCTGCCCATCACTCGATGCAATGG | 21495 |
| Qy | 391 | ACATCTTTCCGCAATTGCTGGGGAAGCGGGTTCCCGTGTGCGAGGTCAAGTTCTGCACGACAGT | 450 |
| Db | 21496 | ATGTCAAGGAGATTGCTCAGGAGCGGGATACCAGGTCAAGTCAAGTTGGCGCGTCAAT | 21555 |
| Qy | 451 | GGTCTCAGAATGGTGATGAGACGTGTTTCCATCATTTGTTGCTCCCAAGG----- | 503 |
| Db | 21556 | GGTCCCAAGAATGGTCGGTTCGATGCCATCTTCCATCACTTCCAAACCGCCCAAGGAGGTG | 21615 |
| Qy | 504 | --CGTACTCTGTCFAACTTTCTCAGGACCATCACTTCGAGGGTCTGATCTCTCTCACCA | 561 |
| Db | 21616 | CTCGCACACTTATTGAGTTCCCGAGCGGATACGAAGCGCGGAATGTGAACACCTTACGA | 21675 |
| Qy | 562 | ATCGACCCCTTCAGCGACTGCAAAACCGTCTGATCGCCATCGAAGTCGGCGAGAGGCTC | 621 |
| Db | 21676 | ACCGTCCCTTGACAGCATTCAAAGCCCGCTCTTGGACCCAGATCGCGAGAGAGTGC | 21735 |
| Qy | 622 | GGTCTTACTTCCATCGTACATGATCCCATCGAAATCGTTGTTCTGCAAGAATGCCTC | 681 |
| Db | 21736 | AGACCTCTCTGCGCCTTACATGATCCCATCGCGCATCATGGCTTGTATCAGATGCGCTG | 21795 |
| Qy | 682 | TCAAAGCCCAATGTTAAGTTGACCGGAGGAACCTCTCGCAGGCGCAAGGTTGTACCGA | 741 |
| Db | 21796 | TCACACAAACGCGAAGATTGACCCAGAGAGCTTGTGCGGAGAGACTATCGTGCCGCCGA | 21855 |
| Qy | 742 | AGCAGCAGACAGCAGCGCGTTTACCAGACATTTCCCATCAGTGAAGTCAAGTCAATCTTT | 801 |
| Db | 21856 | AGCCAGGTCAGCGGCTAC---TCGGGTAGCCCCCGCAATGAGATCGAGGCTATTCTGA | 21912 |
| Qy | 802 | GCGAGAAGCCACTCAGGTGTTTGGCATGAGGTTGACATTACCGATCACTTCTTCAATC | 861 |
| Db | 21913 | GAGACGAATTCGAGGACGTGTCGGAACAGAGAAGTCAGCGTCTGCTGAACCTTCTTTGATC | 21972 |
| Qy | 862 | TCGGTGGACACTCTCTTTGGCCAGGAGCTCAATTTCTCGTATCGACCAACGACTCAAGG | 921 |
| Db | 21973 | TCGGGGGCACTCACTTATGGCCAGAGACTCGCGCCCGCGGTTAGCGCGCGCTTGATG | 22032 |
| Qy | 922 | TCGGTCACTGTCGAAGATGCTTTTGACCATCTCTGTATTTGCGGAATCTAGCATCTGTCA | 981 |
| Db | 22033 | CCCATATTTCCATCAAGATGCTTTGATCAGCCGGTCTGCGCGGATCTTTCGCGCTCCA | 22092 |
| Qy | 982 | TCC 984 | |
| Db | 22093 | TCC 22095 | |

RESULT 4
AAQ54389
ID AAQ54389 standard; DNA; 1713 BP.
XX
XX AC AAQ54389;
XX AC
XX DT 08-JUL-1994 (first entry)
XX

| | |
|----|---|
| DE | T. vasinfecta cyclosporin synthetase gene fragment. |
| XX | |
| KW | Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum; |
| XX | T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase; ss. |
| XX | |
| OS | Tolypocladium vasinfecta. |
| XX | |
| PN | EP578616-A. |
| XX | |
| PD | 12-JAN-1994. |
| XX | |
| PF | 05-JUL-1993; 93EP-0810474. |
| XX | |
| PR | 09-JUL-1992; 92AT-0001403. |
| PR | 08-MAR-1993; 93AT-0000437. |
| PR | 29-APR-1993; 93CH-0001310. |
| PR | 04-MAY-1993; 93CH-0001375. |
| XX | |
| PA | (SANO) SANDOZ LTD. |
| PA | (SANO) SANDOZ PATENT GMBH. |
| PA | (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH. |
| XX | |
| PI | Leitner E, Schneider E, Schoergendorfer K, Weber G; |
| XX | |
| DR | WPI; 1994-010432/02. |
| XX | |
| PT | Isolated DNA sequence - which codes for enzyme having cyclosporin |
| PT | synthetase like activity |
| XX | |
| PS | Example 16; Page 85; 93pp; English. |
| XX | |
| CC | This sequence represents a fragment of the cyclosporin synthetase |
| CC | gene from T. vasinfecta. This fragment was isolated using primers |
| CC | derived from the Tolypocladium niveum (formerly known as T. inflatum |
| CC | GAMS) cyclosporin synthetase gene, corresponding to bases 40309-40328 |
| CC | and 42018-41999. The T. niveum gene encodes an enzyme which catalyses |
| CC | the peptide biosynthesis of cyclosporins and structurally related |
| CC | molecules. The T. niveum sequence may be used for the production of |
| CC | cyclosporin by transforming a vector containing the gene sequence in |
| CC | to a recombinant host. This allows effective production of antibiotic |
| CC | cyclosporin or its derivatives. |
| XX | |
| SQ | Sequence 1713 BP; 390 A; 490 C; 469 G; 364 T; 0 other; |
| XX | |

| Query Match | 18.7%; | Score 187.6; | DB 15; | Length 1713; |
|-----------------------|--------|--|--------|------------------------------------|
| Best Local Similarity | 59.3%; | Pred. No. 7.6e-50; | | |
| Matches | 395; | Conservative | 0; | Mismatches 259; Indels 12; Gaps 4; |
| QY | 338 | GCTATCAACATTCGGTCCAGCCGCGAGGCGACTCATCATCTATCCGTTTCCGACATCTT | 397 | |
| Db | 992 | GATATCGGCACATCAATACGGCGGAAGGAATGCCCTGCTCTCTCAGTGCCGACCTGAT | 1051 | |
| QY | 398 | TCGCATTGCTGGGGAAGCCGGTTCCGTCGAGGTCAGTTTCGACAGCAGATGGTCTCA | 457 | |
| Db | 1052 | TGAGATTGGTTAAGGGATCGGCTTCCAAAGTTGAGACACAGCTGGGCTCGACAACACTCCCA | 1111 | |
| QY | 458 | GAATGTTGATTTGGACGCTGTTTTTCATCATTTGTTTGGTCCCAA-----GGGCGTAC | 508 | |
| Db | 1112 | GGCGCGGCACTCGATGCTGTTTTTCAACCGATTTCGAAAACCAAGACACTCGGGTCTATGT | 1171 | |
| QY | 509 | TCGTGTTCAACTTTCCTACGGACCATCACCTTCGAGGGTCTGATCTC-CTCACCAATCGAC | 567 | |
| Db | 1172 | CATGTTTCAGTTTCCCAACTGAACACAAGGGCGGCTTCGAGCAGTCTCCAGAAATGCC | 1231 | |
| QY | 568 | CCCTTCAGCGACTTGCAAAACCGTCGATCGCCATCGCAAGTCCCGAGAGGCTTCGGTCCCT | 627 | |
| Db | 1232 | CGCTACACCTGGTTTCAGAGCCCGCGCTGGAGCAAAAGGTCGCGAGCGGCTCCCAATCGC | 1291 | |
| QY | 628 | TACTTCCATCGTACATGATGCCATCGAACAATCGTTGTTGGACAAGATGGCTCTCAACG | 687 | |
| Db | 1292 | TGCTTCCATCGTACATGATTTTCCCTCTCGGATCATGTTGCTTCGATCAGATGCCCTCTCACGT | 1351 | |
| QY | 688 | CCAATGGTAAAGTTGACCCGGAAGGAACCTCTCTCGCAGGGCAAAAGTTGTACCGA-AGCAG | 746 | |

Db 1352 CCAACGGCAAGGTGGATCGAAGAGAGTCTGTCACAAAGCCCGGTCTATCCCAACAAATTG 1411
 QY 747 CAGACAGCAGCGCGTTTACCACATTTTCCCATCAGTGAAGTTCGAAGTC-ATTCTTTGGGA 805
 Db 1412 CCGAAGCAGCTTGGACTTTGTGGCGCCAGCAGCGAATCGAGGTCTCTCTGCGA 1471
 QY 806 AGAGCCACTGAGGTGTTTGGCATGAAGGTTGACATTACCGATCACTTCTTCAATCTCGG 865
 Db 1472 AGAATTTACCGATCTACTAGGCGTCAAGGTGCGGATTTACAGACAACTTCTTCGAGTTGGG 1531
 QY 866 TGGACACTCTCTTGGCCACGAAAGCTCATTTCTCGTATCGAACCAAGCTCAAGGTCCG 925
 Db 1532 CGGCGATTCGCTGTGGCCAGAACTGAGCGCAGCTCTAAGTCGACAGCTGGAGCGCGG 1591
 QY 926 TATCACTGTCAAGGATGCTTTTGGACATCTCTGTATTTGCGGATCTAGCATCTGTCATCCG 985
 Db 1592 TGTCACTGTGAAGCAGATCTTTGACCAAGCCAGTACTTCTGCTATCTTGTCTTATTCG 1651
 QY 986 TCAAGG 991
 Db 1652 TCAAGG 1657

RESULT 5

ID AAS51470 standard; DNA; 13029 BP.
 AC AAS51470;
 XX
 DT 13-FEB-2002 (first entry)
 DE Pseudomonas aeruginosa DNA for cellular proliferation protein #55.
 KW Antisense; ds; prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 OS Pseudomonas aeruginosa.
 PN WO200170955-A2.
 PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-20727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR P-PSDB; ANU33611.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Claim 27; Seq ID No 4052; 51pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 13029 BP; 1978 A; 4677 C; 4485 G; 1889 T; 0 other;
 Query Match 5.3%; Score 53; DB 23; Length 13029;
 Best Local Similarity 51.9%; Pred No. 9.8e-06;
 Matches 150; Conservative 0; Mismatches 130; Indels 9; Gaps 1;
 QY 613 AGAGGCTTGGTCTTACTTCCATCGTATCATGATCCATCGAACATCGTTGTTCGGACA 672
 Db 12587 AGCAACTGCGCGCCGACCTGCGGACTACATGGTGGCTGCTGCTGGACC 12646
 QY 673 AGATGCTCTCAACGCCAATGTAAGTTGACCGGAAGAACTCTCTCGCAGGCGAAGG 732
 Db 12647 GGATGCCGCTCAACGCCAAGCTCGACCGCAAGGCGCTGCCGCGTGGACATCG 12706
 QY 733 TTGTACCGAAGCAGCAGACAGCAGCGCGTTTACCACATTTTCCCATCAGTGAAGTTCGAAG 792
 Db 12707 GCCAGATCGAAGAACCGAGCC-----TACCAGGCCCGCGCAACGAACTGGAGGAAA 12757
 QY 793 TCATTCTTTTGGGAAGACCACTGAGTGTGTTGGCATGAAGTTGACATTTACCGATCACT 852
 Db 12758 CCTTGGCGGATCTGCGCGAGGTGCTGAAGTTCGAGCGGTGTTTCGACAACT 12817
 QY 853 TCTTCAATCTCGTGGACACTCTCTCTTGGCCACGAAGCTCATTTCTCG 901
 Db 12818 TCTTGAACCTCGCGGGCATTCGCTGCTGGCCACCCAGATCGCTCGCG 12866

RESULT 6

AAI66165
 ID AAI66165 standard; DNA; 41599 BP.
 AC AAI66165;
 XX
 DT 15-JAN-2002 (first entry)
 DE Bacillus subtilis Iturin A.
 XX
 KW Bacillus subtilis; Iturin A; cation channel; ds.
 XX
 OS Bacillus subtilis.
 XX
 PN JP2001231561-A.
 XX
 PD 28-AUG-2001.
 XX
 PF 18-FEB-2000; 2000JP-0040825.
 XX
 PR 18-FEB-2000; 2000JP-0040825.
 XX
 PA (MASA/) MASADA M.
 XX
 DR WPI; 2001-642167/74.

XX
 PT A gene encoding Iturin A for the production of large amounts of Iturin
 PT A -
 XX
 PS Disclosure; Page 4-17; 18pp; Japanese.
 XX
 CC The invention relates to a gene encoding Iturin A encoding a protein
 CC having an activity of promoting the transfer of a cation to the exterior

CC of the cell through the cation channel of the cell and a promoter for
CC transferring the cation to the exterior of the cell containing Iturin A
CC as the active component. The gene can be used for the preparation of
CC Iturin A in a large quantity.
XX
SQ Sequence 41599 BP; 12054 A; 9173 C; 10241 G; 10131 T; 0 other;
Query Match 5.0%; Score 50; DB 22; Length 41599;
Best Local Similarity 49.7%; Pred. No. 0.00017;
Matches 187; Conservative 0; Mismatches 180; Indels 9; Gaps 2;
QY 599 CATCGAAGTCGCGAGAGGCTTCGCTTACTTCCATCGTACATGCCATCGAAT 658
DB 28370 CAGTGAAGTCGCGAAGAGTGGCCGCGCATTTACCGGATATATGATCCCGCTCATTT 28429
QY 659 GGTGTGTTCTGACAGAGTCTCTCAAGCCAAAGTGAAGTTGACCGGAAGAACTCTC 718
DB 28430 TGTCCAGCTGATGAAGATGCGCT-GACGCCAAAGCGGAAGTTGAATCGCCAGTTATTC 28488
QY 719 TCGCAGGCGAAAGTTGTACCGAAAGCAGCAGCAGCGCGTTTACCGACATTTCCCAT 778
DB 28489 CGGCTCGGTCAAAAGCGTGACAGCGCATAGAGTACGTCGCCCGCAACATTCGCGAG 28548
QY 779 CAGTGAGTGAAGTCTATCTTTTCGGAAGCAGCACTAGAGTGTTCGATGAAGTTGA 838
DB 28549 AAATCCAGCTGACAGCAATTTGGGAGGA-----TGTCCTTGACTAGAGCAGGTGG 28600
QY 839 CATFACCGATCACTTCTTCAATCGCTGACACTCTCTTGGCCACGAGCTCATTT 898
DB 28601 GATCAGAGATCACTTTTGTGATCGCGGCGACACTCCCTCGCGGCAACGCGCTGATTGC 28660
QY 899 TCGTATCGACCAAGCACTCAAGTCCGTATCATCTGTCAAGGATGCTCTTTGACCATCTGT 958
DB 28661 AAAAATACAAAGCAATGATGTCACAAATTCCTTTTCGGGACGCTCTTCGTTCCCAAC 28720
QY 959 ATTTGCGGATCTAGCA 974
DB 28721 CATTTGAACAGCTCGCA 28736

RESULT 7
AAV69559
ID AAV69559 standard; DNA; 1200 BP.
XX
AC AAV69559;
XX
DT 15-MAR-1999 (first entry)
XX
DE Soil derived peptide synthase clone ps7 DNA.
XX
KW Peptide synthase; soil; lichen; antibiotic biosynthesis; humus;
KW therapeutic; immunosuppressor; antitumour agent; pathogen;
KW genetic diversity; ss.
XX
OS Unknown.
XX
PN WO9853097-A2.
XX
PD 26-NOV-1998.
XX
XX 21-MAY-1998; 98WO-CA00488.
XX
XX 22-MAY-1997; 97US-0861774.
XX
PA (TERR-) TERRAGEN DIVERSITY INC.
PI Miao VPW, Seow KT, Waters B, Yap WH;
XX
XX WPI; 1999-070158/06.
DR P-PSDB; AAW82711.
XX
PT New degenerate primers - used for recovering antibiotic biosynthetic
PT DNA from soil/lichen material

XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Claim 17; Page 82-83; 98pp; English.

This sequence encodes a peptide synthase clone, isolated from soil.
This protein is used in a method for the recovery of antibiotic
biosynthetic DNA from humic materials or lichen. The PCR products of the
invention have the potential to be used as therapeutic molecules
including antibiotics, immunosuppressors and antitumour agents. The
method allows access to the reservoir of genetic diversity in soil
pathogenic micro-organisms, in order to find new antibiotics. It also
allows access to novel biosynthetic genes/enzymes that can be used to
produce antibiotics or produce specific compounds, enzymatically,
in vitro.

Sequence 1200 BP; 197 A; 392 C; 384 G; 226 T; 1 other;

Query Match 5.0%; Score 49.6; DB 20; Length 1200;
Best Local Similarity 63.3%; Pred. No. 3.3e-05;
Matches 76; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 596 CGCATCGAAGTCGCGAGAGGCTTCGCTTACTTCCATCGTACATGCCATCGAA 655
DB 928 CGACTCGAGCTGCGCGAGTTCCTCGGACGTTACTCCCCGCGGATGATACCTCGGC 987
QY 656 CATCGTTGTTCTGACAGAGTGCCTCTCAAGCCCAAGTGAAGTTGACCGAAGAACT 715
DB 988 ATTCGTTGTGTGGAGACGCTCCCACTGACCCACACGGAAGGTGGACGAGAGCCCT 1047

RESULT 8

AAAS9145
ID AAAS9145 standard; DNA; 7178 BP.

XX
AC AAAS9145;

XX
DT 07-NOV-2000 (first entry)

XX
DE DNA encoding a peptide synthetase unit-PKS module.

XX
KW Polyketide; antibiotic Tel-Aviv; cell wall synthesis;
KW lipid-disaccharide-pentapeptide; gingivitis; ss.

XX
OS Myxococcus xanthus.

XX
PN EP1026248-A2.

XX
PD 09-AUG-2000.

XX
PF 31-JAN-2000; 2000EP-0300747.

XX
XX 29-JAN-1999; 99US-0240537.

XX
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

XX
PI Rosenberg E, Ron E, Orr E, Paitan Y;

XX
DR WPI; 2000-500254/45.

XX
PT Novel DNA sequence involved in polyketide antibiotic Tel-Aviv
PT production useful for inhibiting cell wall synthesis and in wide range
PT of clinical applications such as treating gingivitis

XX
PS Claim 5; Page 6-8; 66pp; English.

XX
CC The specification describes a DNA sequence which partially encodes
CC a functional portion of polypeptide component required for synthesizing
CC the polyketide antibiotic Tel-Aviv, postmodification of antibiotic
CC Tel-Aviv, or regulation of biosynthesis of antibiotic Tel-Aviv. The
CC antibiotic Tel-Aviv is a macrocyclic polyketide synthesised through
CC the incorporation of acetate, methionine, and glycine. It inhibits cell
CC wall synthesis by interfering with the polymerisation of the
CC lipid-disaccharide-pentapeptide. Antibiotic Tel-Aviv genes are useful
CC in combinatorial genetics, and for encoding protein components for the

CC synthesis, modification and regulation of antibiotic Tel-aviv.
CC Antibiotic Tel-aviv is useful in a wide range of clinical applications
CC such as treating gingivitis. Antibiotic Tel-aviv is also useful for
CC generating new biological agents from its secondary metabolites. The
CC present sequence encodes a protein involved in synthesis of antibiotic
CC Tel-aviv.
XX

SQ Sequence 7178 BP; 1119 A; 2280 C; 2548 G; 1231 T; 0 other;

Query Match 4.9%; Score 49.2; DB 21; Length 7178;
Best Local Similarity 50.9%; Pred. No. 0.00012; Mismatches 158; Indels 9; Gaps 2;
Matches 173; Conservative 0;

QY 607 TCCGCGAGAGGCTTCGGTCTTACCTTCATCGTACATGCCATCGAACATCGTGTGTC 666
Db 2621 TCCGTCACACCTCGCGAAGTTCCTCCGACTACATGTCCTCCGCGACGCTTCGCGG 2680

QY 667 TGGACAAGATGCTTCAACGCCATGTAAAGTTGACCGGAAGAACTCTCTCCGAGG 726
Db 2681 TGGATGCGATTCGCTGTGGGCAATGCAAGTGGACCGGGCCAGCTGATGGCCAGC 2740

QY 727 CAAGGTTGTACCGAAGCAGCAGCAGCGCCGTTACCGACATTTCCCATCAGTGAGG 786
Db 2741 C---GGTGTACCCGCGGAGACATCGCGGTCCATCCGCTGTGTGAGGCA 2797

QY 787 TCGAAGTCAATCTTTCGGAAGCACTGAGGTGTTGGCATGAAGTTGACATTAACG 846
Db 2798 CCTCGTCGAGCTGTGGAAGACGCTCCAGGT-----CAACGAGGTGGTGTGAGG 2851

QY 847 ATCACTTCTTAATCTCGGTGACACTCTCTTGGCCACGAACTCATTTCTGTAATCG 906
Db 2852 ATCGTCTTCTGAAGTGGGGGGACTCGTCTGGCCGCTGCTGGTGGAGGATGA 2911

QY 907 ACCAAGACTCAAGTCCGTATCACTGTCAAGGATGCTT 946
Db 2912 ACCGCGCTTCACAGCGGCTCGCGTCACGACCTGTT 2951

RESULT 9

AA11992

ID AA11992 standard; DNA; 37856 BP.

XX AC AA11992;

XX DT 07-AUG-2000 (first entry)

XX DE S. cellulosum DNA encoding polyketide and heteropolyketide enzymes.

XX KW Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;

XX KW epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;

XX XX plant-protection; ds.

XX OS Sorangium cellulosum.

XX FH Location/Qualifiers

FT CDS complement (3398..6100)

FT FT /tag= a

FT FT /product= "ORF1-trna synthetase"

FT FT /note= "gtg start codon"

FT FT complement (6374..7111)

FT FT /tag= b

FT FT /product= "ORF2-monoxygenase"

FT FT complement (8433..9550)

FT FT /tag= c

FT FT /product= "ORF3-aminotransferase"

FT FT /note= "AGT start codon given in the specification"

FT FT 9855..11393

FT FT /tag= d

FT FT /product= "ORF4- tyrosine/DOPA-Decarboxylase"

FT FT /note= "GTG start codon"

FT FT 12212..13658

FT FT /tag= e

FT FT /product= "ORF5-3-oxoacyl-ACP-reductase"

FT CDS /note= "ACC start codon"
15374..19984
FT FT /tag= f
FT FT /product= "ORF6-polyketide synthase"
20003..27889
FT FT /tag= g
FT FT /product= "ORF7-peptide synthetase"
28251..29400
FT FT /tag= h
FT FT /product= "ORF8-transpeptidase"
complement (30040..31720)
FT FT /tag= i
FT FT /product= "ORF9-regulation element"
31982..32932
FT FT /note= "CGC stop codon"
FT FT /tag= j
FT FT /product= "ORF10-transcription regulator"
33128..33613
FT FT /tag= k
FT FT /product= "ORF11-regulation element"
33661..34077
FT FT /tag= l
FT FT /product= "ORF12-regulation element"
complement (35255..35616)
FT FT /tag= m
FT FT /product= "ORF13-transcription regulator"
complement (35730..36242)
FT FT /tag= n
FT FT /product= "ORF14-transcription regulator"
/note= "GTG start codon"

DE19846493-A1.

13-APR-2000.

09-OCT-1998; 98DE-1046493.

09-OCT-1998; 98DE-1046493.

(GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

Beyer S, Mueller R;

WPI; 2000-294101/26.

DNA sequence coding for products involved in the biosynthesis of

polyketide or heteropolyketide compounds, especially epothilone

Claim 3; Page 20-33; 36pp; German.

XX This invention describes a novel DNA sequence (I) whose expression
XX products effect or are involved in the enzymatic biosynthesis,
XX mutasynthesis or partial synthesis of polyketide or heteropolyketide
XX compounds (II). (I) can be inserted into an expression vector and used
XX to transform or transfect prokaryotic or eukaryotic cells with the aim
XX of obtaining strains that produce large amounts of polyketide or
XX heteropolyketide compounds, especially epothilones, which have cytotoxic
XX and/or immunosuppressant and antibiotic and antifungal activities and
XX are useful as plant-protection agents. This sequence represents the DNA
XX sequence isolated from Sorangium cellulosum which is described in the
XX method of the invention.

SQ Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;

Query Match 4.8%; Score 48.2; DB 21; Length 37856;

Best Local Similarity 58.9%; Pred. No. 0.00063;

Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 830 GAAGGTGACATTACCGATCATTCTCAATCTCGGTGGACACTCTCTTGGCCACGAA 889

Db 19714 GCAGGTAGGCGCAACACGACGATTCTCCAGCTGGCGGCCATCTGTTGCCACGCA 19773

| | | | |
|----|-------|--|-------|
| QY | 890 | GCTCATTTCTCTATCGACCAAGCTCAAGTCCGTATCACTGTCAGGATGCTTTGA | 949 |
| Db | 19774 | GGTCTGTCTCGCGTCTCGACGCTCAAGTGGGATCTCGTTGCGCGAGTCTTCGA | 19833 |
| QY | 950 | CCATCCTGTATTTCGGGATCT | 970 |
| Db | 19834 | TGCGCGCAGCGTTCGCGAGGCT | 19854 |

WO200040704-A1.

13-JUL-2000.

06-JAN-2000; 2000WO-US00445.

06-JAN-1999: 99US-0115435.

05-FEB-1999; 990US-0118848.
05-JAN-2000: 2000US-0477962

RECC \ UNIV CAL TEORNTA

Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
WPI: 2000-465974/40.
P-PSDB: AAB07580, AAB07581, AAB07582, AAB07583, AAB07584, AAB07585,
AAB07586, AAB07587, AAB07588, AAB07589.

New bleomycin gene cluster components useful for peptide and/or
polyketide metabolites, especially bleomycin, production and for
chemically modifying biological molecules -

Claim 8; Page 137-153; 162pp; English.

The present sequence represents the BLM (Bleomycin) gene cluster,
containing open reading frames (ORFs) 31-40. The proteins encoded
by the gene cluster are useful for producing peptides and/or polyketide
metabolites, especially bleomycin or bleomycin analogues. They are
also useful for chemically modifying biological molecules to produce
branched methyl groups, and for coupling amino acids and fatty
acids. They may be reacted with an apo-carrier protein and coenzyme A
to produce a holo-carrier protein. The BLM gene cluster or catalytic
domains can be used individually or collectively to produce
thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
microbial metabolites. The BLM gene cluster may also be used to produce
sugars.

Sequence 18660 BP; 2365 A; 7239 C; 6510 G; 2546 T; 0 other;

* Query Match 4.8%; Score 48; DB 21; Length 18660;
Best Local Similarity 47.2%; Pred. No. 0.0005;
Matches 184; Conservative 0; Mismatches 200; Indels 6; Gaps 1;

QY 596 CGCATCTGAAGTCGCGAGAGGGTTCGGTGCTTTACTTCCATGCTACATGATCCATCGAA 655
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4698 CCCACCAGCTACGCCGGTTCGCCGAGGGGGGTTCGCCGCCATGTCGCCTGCGC 4757
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 656 CATCGTGTGTTCTGGACAAGATGCCCTCAACGCCAATGGTAAGTTGACCGGAAGAAT 715
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4758 GGTGGTCTGCTCGAGGGCGTGCCTGACCTGACGTCGAACGGAAGCTGGACCGCGCGCT 4817
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 716 CTCTCGAGGGCAAAAGTTGTACCGAAGCAGCAGACAGCGCCGTTACCGACATTTC 775
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4818 GCCGCGCCCGCGCGGGCGAGACCGGAATGGAT-----GTCCGGCTTCGTGGCGCGG 4871
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 776 CATCACTGAGGTCCGAAGTCATTCTTTGCGAAGACCACTGAGGTCTTTGGCATGAAGT 835
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4872 CGACATGTTGGAGGAGGTCTGGCCGAGGTCTGGTCCGCCGTGCTGGCGGTGCACCGGT 4931
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 836 TGACATTACCGATFCACTCTTCAATCTCGGTGGACACTCTCTCTTTGGCCACGAAGTCA 895
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4932 CGGTGTGCACGACGACTTCTTCGAGCTGGCGGGCACTGTTCTGTGTGTCCAGTGAT 4991
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 896 TTCTCTGATCGAACACGACTCAAGTCCGTAFACTGTCAGGATGCTTTTGACCATCC 955
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4992 GACCCGGATACGAAAGCTCTCGCGGTCGAGGTGCGGTCGCGGAGCTGTCGAGCGCG 5051
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 956 TGTATTTGCGGATCTAGCATCTGTCATCCG 985
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5052 GACCGTCGAGGAGCTCGCCCGCGCGTCCG 5081
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
ABK74875

ID ABK74875 standard; DNA; 6465 BP.

XX AC ABK74875;

XX AC ABK74875;

XX DT DT

XX 13-AUG-2002 (first entry)

XX Bacillus licheniformis genomic sequence tag (GST) #2166.

XX Differential gene expression; genomic sequenced tag; GST;

KW altered culture condition; environmental stress;

KW physiological provocation; ds.

| | | | |
|-----------------------|---|--|------|
| QY | 810 | GCACCTGAGGTGTTGGCATGAAGTTGACATTACCGATCATTCTTCAATCTCGGTGGA | 869 |
| Db | 4337 | GTGCTGGCATGGATGGCA-----TCGGCGTCCATGATCACTTCTTGGAGGA | 4390 |
| QY | 870 | CACCTCTCTCTTGGCCACGAAGCTCATTTCTGCTATCGACCAACGACTCAAGGT | 922 |
| Db | 4391 | CATTGCTGCTGCTACGCGACAGATGATCGCCGCGGTGGCGACATGCTCCACGT | 4443 |
| RESULT 14 | | | |
| AAF90033 | | | |
| ID | AAF90033 | standard; DNA: 34071 BP. | |
| XX | | | |
| AC | AAF90033; | | |
| XX | | | |
| DT | 06-AUG-2001 | (first entry) | |
| XX | | | |
| DE | | Nucleotide sequence of cosmid a26g1 (coding strand). | |
| XX | | | |
| KW | | Metabolic pathway operon; polyketide; polyketide antibiotic; ss. | |
| XX | | | |
| OS | | Synthetic. | |
| XX | | | |
| PN | WO200140497-A2. | | |
| XX | | | |
| PD | 07-JUN-2001. | | |
| XX | | | |
| PF | 27-NOV-2000; 2000WO-FR03311. | | |
| XX | | | |
| PR | 29-NOV-1999; 99FR-0015032. | | |
| PR | 07-JUN-2000; 2000US-0209800. | | |
| XX | | | |
| PA | (AVET) AVENTIS PHARMA SA. | | |
| XX | | | |
| PI | Jeannin P, Pernodet J, Guerinneau M, Simonet P, Courtois S; | | |
| PI | Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphille K; | | |
| PI | Frostegard A; | | |
| XX | | | |
| DR | WPI: 2001-374849/39. | | |
| XX | | | |
| PT | Collection of nucleic acids from environmental samples, useful for | | |
| PT | identifying e.g. genes encoding polyketide synthases and derived | | |
| XX | | | |
| PS | Claim 35; Page 300-302; 356pp; French. | | |
| XX | | | |
| CC | The specification describes a method for the preparation of a collection | | |
| CC | of nucleic acids from organisms in a soil sample. The method comprises | | |
| CC | milling a dried sample to produce microparticles; suspending these in | | |
| CC | liquid buffer; extraction of nucleic acids from the microparticle; | | |
| CC | passing nucleic acid-containing solution through a molecular sieve; | | |
| CC | chromatography material; and recovering fractions containing purified | | |
| CC | nucleic acids. The nucleic acids are sources for sequences that encode | | |
| CC | either operons involved in a metabolic pathway (specifically polyketide | | |
| CC | synthesis) or polypeptides, particularly for production of therapeutic | | |
| CC | or agricultural compounds, especially polyketide antibiotics. AAF90034-39 | | |
| CC | represent open reading frames (ORFs) of the coding strand of cosmid | | |
| CC | a26g1, and encode type I polyketide synthases. | | |
| XX | | | |
| QY | Sequence 4615 BP; 827 A; 1477 C; 1429 G; 882 T; 0 other; | | |
| XX | | | |
| Query Match | 4.5%; Score 45.4; DB 22; Length 4615; | | |
| Best Local Similarity | 49.8%; Pred. No. 0.0016; | | |
| Matches | 146; Conservative 0; Mismatches 141; Indels 6; Gaps 1; | | |
| QY | 630 | CTTCCATCTGATGATCCATCGAACATCGTTGTTCTGGACAAGATGCTCTCAAGCC | 689 |
| Db | 4157 | CTGCCGCTACATGATCCGACCGGTGGTGGTCTCCACGAATCGCGTACGCC | 4216 |
| QY | 690 | AATGTTAAAGTTGACCGGAAGAACTCTCTCGCAGGGCAAGAGTTGTACCGAAGCAGCAG | 749 |
| Db | 4217 | AACGGAAAAATCGACCGTAACGCCCTGCCGATCCCGACCGCGCGGACCCACGCC | 4276 |
| QY | 750 | ACAGAGCCCGCTTACCGACATTTCCCATCACTGAGGTGGAAGTCAATTTTGGGAAGA | 809 |
| Db | 4277 | GAAGCATTCACGCCCTCCGGAATACTCCGGTGGAAACAGGTACTCGCCACATTTGGGCGAG | 4336 |

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 4, 2003, 19:04:47 ; Search time 57 Seconds
(without alignments)
5385.674 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000

Perfect score: 1001

Sequence: 1 aactgttttttccactcg.....tccgtcaagggtgggttg 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----------------------|-------------------|
| 1 | 240.2 | 24.0 | 46899 | 1 US-08-471-119A-1 | Sequence 1, Appl |
| 2 | 189.2 | 18.9 | 1713 | 1 US-08-471-119A-4 | Sequence 4, Appl |
| 3 | 49.6 | 5.0 | 1200 | 4 US-08-861-774E-85 | Sequence 85, Appl |
| 4 | 39.4 | 3.9 | 68750 | 3 US-09-335-409-1 | Sequence 1, Appl |
| 5 | 39.4 | 3.9 | 68750 | 4 US-09-568-102-1 | Sequence 1, Appl |
| 6 | 39.4 | 3.9 | 68750 | 4 US-09-567-969-1 | Sequence 1, Appl |
| 7 | 39.4 | 3.9 | 68750 | 4 US-09-568-480-1 | Sequence 1, Appl |
| 8 | 39.4 | 3.9 | 68750 | 4 US-09-568-486-1 | Sequence 1, Appl |
| 9 | 39.4 | 3.9 | 68750 | 4 US-09-568-472-1 | Sequence 1, Appl |
| 10 | 39.4 | 3.9 | 68750 | 4 US-09-567-899-1 | Sequence 1, Appl |
| 11 | 39.4 | 3.9 | 71989 | 4 US-09-443-501A-2 | Sequence 2, Appl |
| 12 | 39.2 | 3.9 | 1172 | 4 US-08-861-774E-17 | Sequence 17, Appl |
| 13 | 38.8 | 3.9 | 1178 | 4 US-08-861-774E-91 | Sequence 91, Appl |
| 14 | 38.8 | 3.9 | 7218 | 1 US-08-232-463-14 | Sequence 14, Appl |
| 15 | 37.6 | 3.8 | 4403765 | 4 US-09-103-840A-2 | Sequence 2, Appl |
| 16 | 37.6 | 3.8 | 4411529 | 4 US-09-103-840A-1 | Sequence 2, Appl |
| 17 | 36.4 | 3.6 | 11601 | 2 US-08-222-617A-3 | Sequence 3, Appl |
| 18 | 36.4 | 3.6 | 11601 | 2 US-08-222-617A-24 | Sequence 24, Appl |
| 19 | 35.8 | 3.6 | 7215 | 4 US-09-134-001C-627 | Sequence 627, App |
| 20 | 34.6 | 3.5 | 1177 | 4 US-08-861-774E-23 | Sequence 23, Appl |
| 21 | 33.4 | 3.3 | 4220 | 4 US-09-183-846A-11 | Sequence 11, Appl |
| 22 | 33.4 | 3.3 | 4220 | 4 US-08-961-578C-11 | Sequence 11, Appl |
| 23 | 33.2 | 3.3 | 527 | 1 US-07-998-973A-22 | Sequence 22, Appl |
| 24 | 33.2 | 3.3 | 527 | 2 US-08-452-800-22 | Sequence 22, Appl |
| 25 | 33.2 | 3.3 | 527 | 5 PCT-US92-11353-22 | Sequence 22, Appl |
| 26 | 33.2 | 3.3 | 2620 | 1 US-08-056-200-109 | Sequence 109, App |
| 27 | 33.2 | 3.3 | 2620 | 2 US-08-800-644-109 | Sequence 109, App |

Sequence 93, Appl
Sequence 167, App
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 22, App
Sequence 83, Appl
Sequence 6, Appl
Sequence 81, Appl
Sequence 2065, Ap
Sequence 64, Appl
Sequence 22, App
Sequence 166, App
Sequence 26, Appl
Sequence 1, Appl
Sequence 72, Appl
Sequence 1, Appl
Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-471-119A-1
; Sequence 1, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
US-08-471-119A-1

Query Match 24.0%; Score 240.2; DB 1; Length 46899;
Best Local Similarity 55.0%; Pred. No. 1.1e-69;

| | Matches | 563; | Conservative | 0; | Mismatches | 418; | Indels | 42; | Gaps | 3; |
|----|---------|-------|--|---|--|--------------------------------|--------|-----|------|----|
| QY | 1 | AAC | TGCTTTCTT | CACTCGT | TGAAAGACAGAGG | TTTCCAGGCTCTGGTGGAAACATGTTGAGA | 60 | | | |
| Db | 21076 | ACC | TGCATCTTT | TACATCTT | TTCGCTACGCGCTTGGTGAGAAGATCAAGCAGCTCGAA | 21135 | | | | |
| QY | 61 | TC | TGCGCAAGACAT | GTGAAGCTGTCAANT | GAGCTCAGTGCCTGATCGATATGCCGCTGTG | 120 | | | | |
| Db | 21136 | TT | CTCCCAAGACCAT | TGAAGGCTTACCAAGCAGCTCAGCAAGTACCGATATGCCGAGTAC | 21195 | | | | | |
| QY | 121 | TG | CACGTTCTCGGGT | TTCCACTTTGAGATGAGCTTGTGCTTCCGCTTGAGAAAGATGACTGA | 180 | | | | | |
| Db | 21196 | TAC | ATGTGCGTGCCTCGAGAGAACATCAAC | TATACCAAGTCTCTCCCAAGCCTCGA | 21255 | | | | | |
| QY | 181 | TC | GACITTTCAAGCGAATCAAT | TGAACCAAGTCACTGGTGACCTTCTCAAGTCTTCAG | 240 | | | | | |
| Db | 21256 | TAG | ACTTTGGCGCAGACGGCTTCGACGCGCAGACCCCTATCAAC | TTGCTGAAGGACACA | 21315 | | | | | |
| QY | 241 | AT | GCTGCTATCATGCGAGTCA | GCAMAATTCCTTTGGAATACAGCCCTTTGAAGACAGG | 300 | | | | | |
| Db | 21316 | AG | ATCGCGGACCGTTCGCTATCGGTATATCCCGTACAGCAAGACCATTTGTGAGCGGT | 21375 | | | | | | |
| QY | 301 | TC | GTGCGTTCCCTCAANTAGCAACATCGATGAGTGGCA | ----- | 337 | | | | | |
| Db | 21376 | TT | GTCAACAAAGTCACTGACGAGGATGATATGGAGAAAGCCAGAACTCACTGGACGGAT | 21435 | | | | | | |
| QY | 338 | ----- | GC | TATCAACCATTCGCTCCAGCGCGAGGGGACTCATCACTATCCGTTCCCG | 390 | | | | | |
| Db | 21436 | CAG | CTTGGTTCGACCGTCCCGATGGCCGTCAAAGTGCCCATCACTCATGATGCAATGG | 21495 | | | | | | |
| QY | 391 | AC | ATCTTTTCGCATTCGTGGGGAAGCCGGTTCGCTGCGAGTCAGTTC | TCGCACGACAGT | 450 | | | | | |
| Db | 21496 | AT | GTCAAGGAGATTTGCTCAGGAGCGGGATACCAAGTCAAGTCAAGTGGCGCGTCAAT | 21555 | | | | | | |
| QY | 451 | GG | CTCAGAAATGGTGCATTTGGACGCTGTTTCCATCATTTGTTGCCAAGG | ----- | 503 | | | | | |
| Db | 21556 | GG | TCCAGAATGGTGGCTTCGATGCGCATCTTCCATCACTTGAACCGGCCAAGGAGGTG | 21615 | | | | | | |
| QY | 504 | -- | CGTACTCTGCTCAACTTTCCTACGGACCATCACTTTCGAGGGTCTGATCTCCTCACCA | 561 | | | | | | |
| Db | 21616 | CT | CGCACACTTATTAGTTTCCGACGGATTACGAAGCCGGAATGTGNAACACTTAA | 21675 | | | | | | |
| QY | 562 | AT | CGACCCCTTCAGGACTGCAAAACCGTGTGATGCGCATCGAATCGCGAGAGGCTC | 621 | | | | | | |
| Db | 21676 | AC | GTCCCTGTAAACAGCATTCAAAGCCGCCGTCTTGGACGACAGATCGCGAGAAGCTGC | 21735 | | | | | | |
| QY | 622 | GG | TCCTTACTTCCATCGTACATGATCCCATCGAACATCGTTGTTCTGACAAAGATGCCTC | 681 | | | | | | |
| Db | 21736 | AG | ACCTCTCTCGCGCTTACATGATCCCATCGCGATCATGGTCTTGATCAGATGGCTG | 21795 | | | | | | |
| QY | 682 | TC | AAGCCCAATGGTAAAGTTGACCGGAAGGAACCTCTCTCGAGGGGCCAAAGGTTGTACCGA | 741 | | | | | | |
| Db | 21796 | TC | AAACAACAACGGCAAGATTGACCCCAAGGAGCTTGTGCGGAGAGCTATGCTGGGCCCGA | 21855 | | | | | | |
| QY | 742 | AG | CAGCAGACACAGCGCGGTTACCGACATTTCCCATCAGTGAAGTCAATCTTTT | 801 | | | | | | |
| Db | 21856 | AG | CAAGGTACGCGCTAC-- | ---TGGGTAGCCCCCGCAATAGATCAGAGCTATTCTGA | 21912 | | | | | |
| QY | 802 | GC | GAAGAAGCCACTGAGTGTGTCATGAAGTTGACATTAACCGATCACTTCTTCAATC | 861 | | | | | | |
| Db | 21913 | GAG | ACGAATTCGAGGACGTGCTCGGAACAGAAATCAGGTCTCGGATAACTTCTTTGATC | 21972 | | | | | | |
| QY | 862 | TC | GGTGGACACTCTCTTTGGCCACGAAGCTCATTTCTCGTATGACCAACGACTCAAGG | 921 | | | | | | |
| Db | 21973 | TC | GGCGGCACTCACTTATGCGCACGAAGCTTCGCGCGCCGCTTAGCGCGCCTTGATG | 22032 | | | | | | |
| QY | 922 | TC | GTCTCACTGTCAAGGATGCTTTTGACCATCCGTCGATTTTGGGATCTAGCATCTGTCA | 981 | | | | | | |
| Db | 22033 | CC | ATATTTCCATCAAGATGCTTTGATCAGCCGGTGTGGGGGATCTTTGCGCGCTCCA | 22092 | | | | | | |
| QY | 982 | TCC | 984 | | | | | | | |
| Db | 22093 | TCC | 22095 | | | | | | | |

RESULT 2
US-08-471-119A-4
Sequence 4, Application US/08471119A
Patent No. 5827706
GENERAL INFORMATION:
APPLICANT: Leitner, Ernst
APPLICANT: Schneider, Elisabeth
APPLICANT: Schoergendorfer, Kurt
APPLICANT: Weber, Gerhard
TITLE OF INVENTION: Cyclosporin Synthetase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5827706artis Corporation
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,119A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kassenoff, Melvyn
REGISTRATION NUMBER: 26,389
REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 503 8474
TELEFAX: 201 503 8807
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1713 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Necosmospora vasinfecta
US-08-471-119A-4

| | Query Match | 18.9%; | Score 189.2; | DB 1; | Length 1713; |
|----|-----------------------|--|--------------------|------------|--------------|
| | Best Local Similarity | 59.5%; | Pred. No. 2.2e-53; | | |
| | Matches 396; | Conservative 0; | Mismatches 258; | Indels 12; | Gaps 4; |
| QY | 338 | GCTATCAACCATTCGGTCCAGCCGCGAGGCGACTCATCACTATPCCGTTCCCGACATCTT | 397 | | |
| Db | | | | | |
| QY | 992 | GATATCGGCACCTCAATCACGGCGAAGGAATGCCCTGCTCTCTCAGTGCCGACCTGAT | 1051 | | |
| Db | | | | | |
| QY | 398 | TGCGATTGTGGGGAAGCGGGTTCGTCGTCGAGGTCAGTCTTCGACGACAGTGGTCTCA | 457 | | |
| Db | | | | | |
| QY | 1052 | TGAGATTGTAGGGGATCGGCTTCCAAGTTAGACCAAGCTGGGGCTCGACAACACTCCCA | 1111 | | |
| Db | | | | | |
| QY | 458 | GAATGGTCATTTGGACGCTGTTTTTCCATCATTTGCTGCCAA-----GGCGCTAC | 508 | | |
| Db | | | | | |
| QY | 1112 | GCGCGCGGACTCGATGCTGTTTTTCACCAGATTCGAAAACCAAGACACTCGGGTCA GT | 1171 | | |
| Db | | | | | |
| QY | 509 | TCGTGTCAACTTTCCCTACGGACCAATCACTTTCGAGGGTCTGATCTC--CTCACCAAATCGAC | 567 | | |
| Db | | | | | |
| QY | 1172 | CATGTTTCAGGTTCCCAACTGAACACAAAGGGCGCGTCTTCGAGCAGTCTCACGAATCGCC | 1231 | | |
| Db | | | | | |
| QY | 568 | CCCTTCAGCGACTGCAAAACCGTCTGATCGCCATCGAAGTCCCGAGAGCGCTTCGGTCCCT | 627 | | |
| Db | | | | | |
| QY | 1232 | CGCTACACCTGGTTACAGAGCCCGCGCTCGAGSCAAAGGTCGCGAGGCGCTGCAATCGC | 1291 | | |
| Db | | | | | |

```

QY 628 TACTTCCATCGTACATGATCCATCGAAGATCGTCTCTCAAG 687
    |||||
Db 1292 TCGTCCATCGTACATGATTCCTCTCGGATCATGTTGCTCGATCAGATCGCTCAAGT 1351
    |||||
QY 688 CCAATGGTAAAGTTGACCGAGGAACTCTCTCGCAGGCAAGGTTGTACCGA-AGCAG 746
    |||||
Db 1352 CCAAGCGGAAGTGTGATCGCAAGAAGTCTCGCAGCAAGCCCGGTCAATCCCAACAATG 1411
    |||||
QY 747 CAGACAGACGCGCGTTACCGACATTTCCCATCAGTGAGTCCGAAGTC-ATTCTTTGCGA 805
    |||||
Db 1412 CGCGAAGCAGCTTGACATTTGTCGCGCCGACCGCAAGTTCGATCGGTTCTCTGCGA 1471
    |||||
QY 806 AGAAGCCATGAGGTTTGGCATGAAGTTGACATTCACGATCACTTCTTCAATCTCGG 865
    |||||
Db 1472 AGAATTTACCGATCTACTAGGCGTCAAGTTCGCGCAAGTTCACAGCAACTTCTTCAGTTGG 1531
    |||||
QY 866 TGGACACTCTCTTTGGCCACCAAGCTCATTTCTCGTATCGTACGACCAAGTCAAGTCCG 925
    |||||
Db 1532 CGGCCATTCGCTGCTGGCCACCAAGTTCGATCGTATTCGATCGTATTCGATTCG 1591
    |||||
QY 926 TATCACTGTCAAGGATGCTTTGACCATCCTGTATTTGGGATCTAGCATCTGTCTATCG 985
    |||||
Db 1592 TGTCACTGTGAAGCAGATCTTTGACCGCAAGTTCGATCTGTCTATTCGATTCG 1651
    |||||
QY 986 TCAAGG 991
    |||||
Db 1652 TCAAGG 1657
    |||||

```

RESULT 3

```

US-08-861-774E-85
; Sequence 85, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Ho, Yip
; APPLICANT: Miao, Vivian
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps30
US-08-861-774E-85

```

```

Query Match 5.0%; Score 49.6; DB 4; Length 1200;
Best Local Similarity 63.3%; Pred. No. 1.8e-06;
Matches 76; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 596 CGCCATCGAAGTCCGCGAGAGGCTTCGGTCTTACTTCCATCGTACATGATCCCATCGAA 655
    |||||
Db 928 CGACTCGAGCTGCGGAGTTCTCTGGGACGTACTCCCGGAGCGGATACCTTCGCG 987
    |||||
QY 656 CATCGTTGTTCGCAAGATGCTCTCAACGCCAATGTAAGTTGACCGGAAGAACT 715
    |||||
Db 988 ATTCTGTGTCTGGAGACGCTCCCACTGACCCACCAAGGAGGTGGACCGAGAGCCCT 1047
    |||||

```

RESULT 4

```

US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James

```

```

; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 3.9%; Score 39.4; DB 3; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGCGAGAGGCTTCGGTCTTACTTCCATCGTACATGCCCATCGA 654
    |||||
Db 15638 TCGCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTCCCTACAG 15697
    |||||
QY 655 ACATCGTTTCTTGGACAAGATGCCCTCTCAACGCCAATGTTAAAGTTGACCGGAAGAAC 714
    |||||
Db 15698 TCTTCGTGGAGCTCGATGCGTTGCCGCTGACGTCCAAACGCAAGTTCGATCGTAAAGGCC 15757
    |||||
QY 715 T 715
    |||||
Db 15758 T 15758
    |||||

```

RESULT 5

```

US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

```

```

Query Match 3.9%; Score 39.4; DB 4; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGCGAGAGGCTTCGGTCTTACTTCCATCGTACATGCCCATCGA 654
    |||||
Db 15638 TCGCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTCCCTACAG 15697
    |||||
QY 655 ACATCGTTTCTTGGACAAGATGCCCTCTCAACGCCAATGTTAAAGTTGACCGGAAGAAC 714
    |||||
Db 15698 TCTTCGTGGAGCTCGATGCGTTGCCGCTGACGTCCAAACGCAAGTTCGATCGTAAAGGCC 15757
    |||||
QY 715 T 715
    |||||
Db 15758 T 15758
    |||||

```

```
RESULT 6
US-09-567-969-1
: Sequence 1, Application US/09567969
: Patent No. 6355457
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,969
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match      3.9%; Score 39.4; DB 4; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGGGAGAGGCTTCGGTCCCTTACTTCCATCGTACATGATCCCATCGA 654
    |||||
Db 15638 TCGCCGATATGCTTCGGGAGCTTCTTGAGGACCAAACTACCCGAGTACATGCTGCTACAG 15697

QY 655 ACATCGTTGTTCTGGACAAGATGCTCTCAACGCCAATGTTAAAGTTGACCGGAAGGAAC 714
    |||||
Db 15698 TCITCGTGGAGCTCGATGCGTTCGCGCTGAGTCCAAACGGCAAGTCTGATCGTAAGGCC 15757

QY 715 T 715
Db 15758 T 15758

RESULT 7
US-09-568-480-1
: Sequence 1, Application US/09568480
: Patent No. 6355458
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,480
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match      3.9%; Score 39.4; DB 4; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGGGAGAGGCTTCGGTCCCTTACTTCCATCGTACATGATCCCATCGA 654
    |||||
Db 15638 TCGCCGATATGCTTCGGGAGCTTCTTGAGGACCAAACTACCCGAGTACATGCTGCTACAG 15697

QY 655 ACATCGTTGTTCTGGACAAGATGCTCTCAACGCCAATGTTAAAGTTGACCGGAAGGAAC 714
    |||||
Db 15698 TCITCGTGGAGCTCGATGCGTTCGCGCTGAGTCCAAACGGCAAGTCTGATCGTAAGGCC 15757

QY 715 T 715
Db 15758 T 15758

RESULT 8
US-09-568-486-1
: Sequence 1, Application US/09568486
: Patent No. 6355459
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,486
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match      3.9%; Score 39.4; DB 4; Length 68750;
Best Local Similarity 57.9%; Pred. No. 0.059;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 595 TCGCCATCGAAGTCCGGGAGAGGCTTCGGTCCCTTACTTCCATCGTACATGATCCCATCGA 654
    |||||
Db 15638 TCGCCGATATGCTTCGGGAGCTTCTTGAGGACCAAACTACCCGAGTACATGCTGCTACAG 15697

QY 655 ACATCGTTGTTCTGGACAAGATGCTCTCAACGCCAATGTTAAAGTTGACCGGAAGGAAC 714
    |||||
Db 15698 TCITCGTGGAGCTCGATGCGTTCGCGCTGAGTCCAAACGGCAAGTCTGATCGTAAGGCC 15757

QY 715 T 715
Db 15758 T 15758

RESULT 9
US-09-568-472-1
: Sequence 1, Application US/09568472
: Patent No. 6358719
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,472
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
```


Db 273 CGCGTCGAGTGGCGGATCATCTTCGACGCGGTTCCCGAGTACATGATTCGCGTGC 214
Qy 656 CATCGTTGTTCTGACAAAGATGCTCTCAACGCAATGTTAAAGTTGACCGGAAGAACT 715
Db 213 ATTCGTGAGTCCGCGAGATGCGCGCATGGCGGAGCGCAAGATCGATCTGAAGTGCT 154
Qy 716 CTCT 719
Db 153 GCCT 150

RESULT 13

US-08-861-774E-91/c
; Sequence 91, Application US/08861774E

; Patent No. 6297007

; GENERAL INFORMATION:

; APPLICANT: Waters, Barbara

; APPLICANT: Miao, Vivian

; APPLICANT: Ho, Yap

; APPLICANT: Tong, Seow

; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR

; TITLE OF INVENTION: BIOACTIVE MOLECULES

; FILE REFERENCE: 9993-006

; CURRENT APPLICATION NUMBER: US/08/861,774E

; CURRENT FILING DATE: 1997-05-22

; NUMBER OF SEQ ID NOS: 94

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 91

; LENGTH: 1178

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Clone ps25

US-08-861-774E-91

Query Match:

Best Local Similarity 3.9%; Score 38.8; DB 4; Length 1178;

Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 602 CGAAGTCCGCGAGAGGCTCGGTCCTTACTTCCATCGTACATGATCCCATCGAACATCGT 661

Db 270 CGAGTCCGCGAGCGGCTCGCGGAGACTCCCGGATGATGTCGACAGGTGATCGT 211

Qy 662 TGTTCTGGACAAGATGCCTCTCAACGCCATGTAAAGTTGACCGGAAGAACT 715

Db 210 ATCGCTCGCGCGCTGCGCTCACGCGGACGACAAAGATCGACCGCAAGGCCCT 157

RESULT 14

US-08-232-463-14

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/232,463

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14

Query Match

Best Local Similarity 3.9%; Score 38.8; DB 1; Length 7218;

Matches 10; Conservative 210; Mismatches 162; Indels 0; Gaps 0;

Qy 301 TCGTCGCTTCCCTCAATAGCAACATCGATGAGTGGCAGCTATCAACCATTCGGTCCAGCG 360

Db 1068 YY 1127

Qy 361 CCGAGGCGAGTCACTATCCGTTCCCGACATCTTCGCATTCCTGGGGAAGCCGGT 420

Db 1128 YY 1187

Qy 421 TCGGTGCGAGTCACTTCTGCAGCAGTGGTCTCAGATGGTGCATTTGACGCTGTTT 480

Db 1188 YY 1247

Qy 481 TCCATCATCTTGTCTCCCAAGGCGTACTCTGGTCAACTTCTTACGAGCATCACCTTC 540

Db 1248 YY 1307

Qy 541 GAGGCTGTGATCTCCTCAACATCGACCCCTTCAGCGACTGCAAAACCGTGTATCGCCA 600

Db 1308 YY 1367

Qy 601 TCGAAGTCCGCGAGAGGCTTCGTCCTTACTTCCATGTCATGATCCCATCGAACATCG 660

Db 1368 YY 1427

Qy 661 TTGTTCTGGACAAGATGCCTCT 682

Db 1428 YYYYYYGTACCAATCTTCT 1449

RESULT 15

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 20:51:29 ; Search time 170 Seconds
(without alignments)
8240.805 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000
Perfect score: 1001
Sequence: 1 aactcgttttccactcg.....tcgtcaaggctgggttg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----|---------------------|
| 1 | 53 | 5.3 | 13029 | 10 | US-09-815-242-4052 |
| 2 | 49.6 | 5.0 | 1200 | 10 | US-09-924-256A-85 |
| 3 | 46.8 | 4.7 | 6465 | 10 | US-09-974-300-2166 |
| 4 | 46.8 | 4.7 | 8268 | 10 | US-09-974-300-2169 |
| 5 | 42.8 | 4.3 | 88421 | 9 | US-09-976-059-1 |
| 6 | 40.4 | 4.0 | 7347 | 10 | US-09-815-242-7773 |
| 7 | 39.8 | 4.0 | 3798 | 10 | US-09-974-300-2175 |
| 8 | 39.8 | 4.0 | 3798 | 10 | US-09-974-300-2193 |
| 9 | 39.4 | 3.9 | 68750 | 9 | US-10-014-717-1 |
| 10 | 39.2 | 3.9 | 1172 | 10 | US-09-924-256A-17 |
| 11 | 38.8 | 3.9 | 1178 | 10 | US-09-924-256A-91 |
| 12 | 38 | 3.8 | 1160 | 9 | US-10-184-634-592 |
| 13 | 37.2 | 3.7 | 499 | 9 | US-10-184-644-592 |
| 14 | 37.2 | 3.7 | 499 | 9 | US-10-184-634-592 |
| 15 | 36.4 | 3.6 | 513 | 9 | US-10-123-155-536 |
| 16 | 36.2 | 3.6 | 547 | 9 | US-10-066-543-1765 |
| 17 | 36.2 | 3.6 | 4541 | 9 | US-10-123-036-3 |
| 18 | 35.2 | 3.6 | 4541 | 10 | US-09-880-107-3785 |
| 19 | 35.6 | 3.6 | 290 | 10 | US-09-294-093B-2985 |

| | | | | | | |
|----|------|-----|-------|----|---------------------|-------------------|
| 20 | 35.6 | 3.6 | 1803 | 9 | US-10-166-087-43 | Sequence 43, Appl |
| 21 | 35.6 | 3.6 | 32539 | 9 | US-10-166-087-1 | Sequence 1, Appl |
| 22 | 35.4 | 3.5 | 3471 | 10 | US-09-974-300-2167 | Sequence 2167, Ap |
| 23 | 35.2 | 3.5 | 270 | 9 | US-10-060-036-2566 | Sequence 2566, Ap |
| 24 | 35.2 | 3.5 | 400 | 7 | US-08-781-986A-3817 | Sequence 3817, Ap |
| 25 | 35.2 | 3.5 | 400 | 7 | US-08-781-986A-3926 | Sequence 3926, Ap |
| 26 | 35.2 | 3.5 | 29555 | 7 | US-08-781-986A-206 | Sequence 206, App |
| 27 | 34.6 | 3.5 | 716 | 9 | US-10-123-155-96 | Sequence 23, Appl |
| 28 | 34.6 | 3.5 | 1177 | 9 | US-09-924-256A-23 | Sequence 346, App |
| 29 | 34 | 3.4 | 671 | 9 | US-10-184-644-346 | Sequence 346, App |
| 30 | 34 | 3.4 | 671 | 9 | US-10-184-634-346 | Sequence 134, App |
| 31 | 33.6 | 3.4 | 518 | 10 | US-09-833-381-134 | Sequence 135, App |
| 32 | 33.6 | 3.4 | 518 | 10 | US-09-833-381-135 | Sequence 191, App |
| 33 | 33.4 | 3.3 | 244 | 10 | US-09-923-876-191 | Sequence 6617, Ap |
| 34 | 33.4 | 3.3 | 410 | 9 | US-09-918-995-6617 | Sequence 1, Appl |
| 35 | 33.4 | 3.3 | 4157 | 7 | US-08-556-422-1 | Sequence 2171, Ap |
| 36 | 33.4 | 3.3 | 7158 | 10 | US-09-974-300-2171 | Sequence 21047, A |
| 37 | 33.2 | 3.3 | 445 | 9 | US-09-918-995-21047 | Sequence 1125, Ap |
| 38 | 33.2 | 3.3 | 487 | 10 | US-09-833-381-1125 | Sequence 1126, Ap |
| 39 | 33.2 | 3.3 | 487 | 10 | US-09-833-381-1126 | Sequence 9039, Ap |
| 40 | 33.2 | 3.3 | 599 | 9 | US-10-198-846-9039 | Sequence 223, App |
| 41 | 33.2 | 3.3 | 2619 | 10 | US-09-962-832-223 | Sequence 93, Appl |
| 42 | 32.8 | 3.3 | 1178 | 10 | US-09-924-256A-93 | Sequence 112, App |
| 43 | 32.6 | 3.3 | 910 | 9 | US-10-123-155-112 | Sequence 433, App |
| 44 | 32.6 | 3.3 | 1188 | 9 | US-09-764-868-433 | Sequence 9546, Ap |
| 45 | 32.6 | 3.3 | 1551 | 10 | US-09-815-242-9546 | |

ALIGNMENTS

RESULT 1

US-09-815-242-4052

Sequence 4052, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlssen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4052

LENGTH: 13029

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-815-242-4052

Query Match 5.3%; Score 53; DB 10; Length 13029;
Best Local Similarity 51.9%; Pred. No. 7.9e-07;

| Matches | 150; | Conservative | 0; | Mismatches | 130; | Indels | 9; | Gaps | 1; |
|---------|-------|---|-------|------------|------|--------|----|------|----|
| QY | 613 | AGAGGTTTCGGTCCCTTACTTCATCGTACATGATCCCATCGAACATCGTTTCTTCTGGACA | 672 | | | | | | |
| Db | 12587 | AGCAACTGGCGCCGACCTGCGGGACTACATGGTCCGCTGCTACTGGCTGGTGTGGACC | 12646 | | | | | | |
| QY | 673 | AGATGGCTCTCAACGCCAATGTAAAGTTGACCGGAAGSAACTCTCTCGAGGGCAAGG | 732 | | | | | | |
| Db | 12647 | GGATGCCGTCTAAGCCGCAACGCAAGCTGACCGCAAGCGGTGCCGGCTGTGCACATCG | 12706 | | | | | | |
| QY | 733 | TTGTACCGAAGCAGCAGACAGCAGCGCGGTACCGACATTTCCCATCAGTCAAGTCTCGAAG | 792 | | | | | | |
| Db | 12707 | GCCAGATGCAGAACCAAGGCC-----TACGAGGCCCGCGCAACAACTGGAGGAAA | 12757 | | | | | | |
| QY | 793 | TCATTCTTTGCGAAGAAGCCACTGAGGTGTTTGGCATCAAGGTGTGACATTACCCATCACT | 852 | | | | | | |
| Db | 12758 | CCCTGGCGGGATCTGGGCCGAGGTCTGAAGGTGCAGCGGTGCGGGGTGTCGACAACT | 12817 | | | | | | |
| QY | 853 | TCCTCAATCTCGGTGACACTCTCTCTTTGGCCACGAAGCTCATTTCTCG | 901 | | | | | | |
| Db | 12818 | TCCTCGAACTCGCGGGGATTCGCTGCTGGGCACCCAGATCGCTCGCG | 12866 | | | | | | |

```

RESULT 2
US-09-924-256A-85
; Sequence 85, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924.256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 85
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; - OTHER INFORMATION: Description of Artificial Sequence: Clone ps30
US-09-924-256A-85

```

[illegible]

RESULT 3
US-09-974-300-2166
; Sequence 2166, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression.
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300

```

; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2166
; LENGTH: 6465
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2166

```

| | Query Match | 4.7% | Score 46.8 | DB 10 | Length 6465 |
|----|-----------------------|---|-------------------|--------|---------------|
| | Best Local Similarity | 63.2% | Pred. No. 7.4e-05 | | |
| | Matches | 72 | Conservative | 0 | Mismatches 42 |
| | | | | Indels | Gaps |
| QY | 602 | CGAAGTCCGGAGAGCGCTTCGGTCCCTACTTCCATCGTACATGATCCATCGAACATCGT | 661 | | |
| | | | | | |
| Db | 3662 | CGAAATGAGAGAGCGCTTTCTGAGCGGCTTCGTCGTACATGATCCGTCATATTTCGT | 3721 | | |
| | | | | | |
| QY | 662 | TGTTCTGGACAAGATGCCCTCTCAACGCCAATGTTAAAGTTGACCGGAAGAACT | 715 | | |
| | | | | | |
| Db | 3722 | ACGGTGTGGATAAAATCCCGCTTGCGCCGAACGAAAAGTGGACCGAGCGCGCT | 3775 | | |
| | | | | | |

```

RESULT 4
US-09-974-300-2169
; Sequence 2169, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2169
; LENGTH: 8268
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2169

```

| | Query Match | 4.7%; | Score 46.8; | DB 10; | Length 8268; |
|------|---|-----------------|--------------------|-----------|--------------|
| | Best Local Similarity | 63.2%; | Pred. No. 8.6e-05; | | |
| | Matches 72; | Conservative 0; | Mismatches 42; | Indels 0; | Gaps 0; |
| 602 | CGAAATCCGCGAGAGCGCTTCGGTCTCTACTTCATCGTATCATGATCCCATCGAACATCGT | | | | |
| 3363 | CGAAATGAGAGAGCGCTCTTCTGAGCGGCTCCGTCGTACATGATCCGTCATATTTCGT | | | | |
| 602 | TGTTCTGCACAGATGCCTCTCAACGCCAATGGTAAAGTTGACCGGAAGAACT | | | | |
| 3423 | AACGGTGGATAAAATTCGCGTTTCGCGCAACCGAAAAGTGGACCGACGGCGCGT | | | | |

```

RESULT 5
US-09-976-059-1
; Sequence 1, Application US/09976059
; Patent No. US20020164747A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfreddo
; TITLE OF INVENTION: Genes and Proteins of Ramoplanin
; FILE REFERENCE: 3019-pCT

```

CURRENT APPLICATION NUMBER: US/09/976,059
CURRENT FILING DATE: 2001-10-15
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 88421
TYPE: DNA
ORGANISM: Actinoplanes sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2077)..(3078)
OTHER INFORMATION: ORF 1; positive strandedness
NAME/KEY: misc_feature
LOCATION: (3118)..(4032)
OTHER INFORMATION: ORF 2; positive strandedness
NAME/KEY: misc_feature
LOCATION: (4038)..(5048)
OTHER INFORMATION: ORF 3; positive strandedness
NAME/KEY: misc_feature
LOCATION: (4038)..(5048)
OTHER INFORMATION: ORF 3; positive strandedness
NAME/KEY: misc_feature
LOCATION: (5048)..(6058)
OTHER INFORMATION: ORF 4; negative strandedness
NAME/KEY: misc_feature
LOCATION: (6058)..(7068)
OTHER INFORMATION: ORF 5; negative strandedness
NAME/KEY: misc_feature
LOCATION: (7068)..(8078)
OTHER INFORMATION: ORF 6; negative strandedness
NAME/KEY: misc_feature
LOCATION: (8078)..(9088)
OTHER INFORMATION: ORF 7; positive strandedness
NAME/KEY: misc_feature
LOCATION: (9088)..(10098)
OTHER INFORMATION: ORF 8; negative strandedness
NAME/KEY: misc_feature
LOCATION: (10098)..(11108)
OTHER INFORMATION: ORF 9; negative strandedness
NAME/KEY: misc_feature
LOCATION: (11108)..(12118)
OTHER INFORMATION: ORF 10; negative strandedness
NAME/KEY: misc_feature
LOCATION: (12118)..(13128)
OTHER INFORMATION: ORF 11; positive strandedness
NAME/KEY: misc_feature
LOCATION: (13128)..(14138)
OTHER INFORMATION: ORF 12; positive strandedness
NAME/KEY: misc_feature
LOCATION: (14138)..(15148)
OTHER INFORMATION: ORF 13; positive strandedness
NAME/KEY: misc_feature
LOCATION: (15148)..(16158)
OTHER INFORMATION: ORF 14; positive strandedness
NAME/KEY: misc_feature
LOCATION: (16158)..(17168)
OTHER INFORMATION: ORF 15; positive strandedness
NAME/KEY: misc_feature
LOCATION: (17168)..(18178)
OTHER INFORMATION: ORF 16; positive strandedness
NAME/KEY: misc_feature
LOCATION: (18178)..(19188)
OTHER INFORMATION: ORF 17; positive strandedness
NAME/KEY: misc_feature
LOCATION: (19188)..(20198)
OTHER INFORMATION: ORF 18; positive strandedness
NAME/KEY: misc_feature
LOCATION: (20198)..(21208)
OTHER INFORMATION: ORF 19; positive strandedness
NAME/KEY: misc_feature
LOCATION: (21208)..(22218)
OTHER INFORMATION: ORF 20; negative strandedness

LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
NAME/KEY: misc_feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
NAME/KEY: misc_feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
NAME/KEY: misc_feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
NAME/KEY: misc_feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
NAME/KEY: misc_feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
NAME/KEY: misc_feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
NAME/KEY: misc_feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
NAME/KEY: misc_feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
NAME/KEY: misc_feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
NAME/KEY: misc_feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
NAME/KEY: misc_feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; positive strandedness
NAME/KEY: misc_feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1

Query Match 4.3%; Score 42.8; DB 9; Length 88421;
Best Local Similarity 61.8%; Pred. No. 0.0084;
Matches 68; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 606 GTCCGCGAGGCTTCCTTACTTCATCGTACATGATCCCATCGAATCGTGTGT 665
Db 55148 GTCCGCGCGACGTGCGCCGCGCTACATGTCCTCGCGCTCGTGTGTGT 55207
QY 666 CTGGACAAGATGCTCTCAACGCCCAATGGTAAAGTTGACCGGAAGAACT 715
Db 55208 CTGCGCGACCTGCGCTGACCGCAAGCTCGACCGCAAGCGCT 55257

RESULT 6

US-09-815-242-7773
Sequence 7773, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of essential genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7773
; LENGTH: 7347
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7347)
; US-09-815-242-7773

```

```

Query Match      4.0%; Score 40.4; DB 10; Length 7347;
Best Local Similarity 53.9%; Pred. No. 0.013;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 607 TCGCGAGAGGCTTCGGTCCCTTACTTCCATCGTACATGCCATCGAAGTCTGTTTC 666
    ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 6089 TCGCGGAGTGCCTGAAGCGGCACCTGCGGACTACATGTTGCGCGGCACCTGATGCTGC 6148

QY 667 TGCACAAAGATCCCTCTCAAGCCAAATGGTAAATTTGACCGGAAGGACATCTTCGCGAGGG 726
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6149 TGGAGCGATGCCGCTGACGGTCAATGGCAAGCTCGACCGGAGGCGTTGCCGCAACCGG 6208

QY 727 CAAAGTTGTACCGAAGCAGCAGACAGCAGCGCC 760
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6209 ATCGAGCTGTGCGCAGCAGCGCTATCGAGCGCC 6242

```

```

RESULT 7
US-09-974-300-2175
; Sequence 2175, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2175
; LENGTH: 3798
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2175

```

```

Query Match      4.0%; Score 39.8; DB 10; Length 3798;
Best Local Similarity 50.7%; Pred. No. 0.015;
Matches 185; Conservative 0; Mismatches 162; Indels 18; Gaps 3;

QY 630 CTTCCATCGTACATGCCATCCATCGACATCGTTGTTCTGGACAAGATGCCCTCTCAAGCC 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2740 CTACCGGCTTACATGGTGGCGCTGCGTATATCATGCTTGAAGAGCTGCCGCTTACCGCT 2799

QY 690 AATGTTAAAGTTGACCGGAAGAACTCTCTCGCAGGGCAAGGTTGTACCGAAGCAGCAG 749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2800 AACGGAAGTCAACCG--ACGCTTCTCCTGAAGCCGATGGCCGCCGAATTCACAG 2857

QY 750 ACAGCAGCGCGGTTTACCGACATTTCCCATCAGTGAAGTCAATTTCTTGGGAAGAA 809
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2858 AGCGCCGTC-----ACCCGCAATCTAGGGAAGAAAGCTTGCAGTGATT 2904

QY 810 GCCACTGAGTGTGTTGCA---TGAAGTTGACATTTACCGATCATCTTCTCAATCTCGGT 866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2905 TGGTCTGAGGTTCTCGGCAGACAGCAGATCGGAATAGACGAGAATTTTTTGAATCGG 2964

QY 867 GGACACTCTCTTGGCCACGAAGCTCATTTCTCGTATCGAACACGACTCAAGTCCGT 926
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2965 GGACATTCATTGAAAGCCATGCGCTCGCGGATGCTGAAAGATCTCGCAATCGAT 3024

QY 927 ATCACTGTCAAGGATGCTTTTGACCATCTGTTTTCGGATCTTAGCATCTGTCTATCG 986
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3025 GTTCCGGTGAATGTGTTTGAATGCCACATCGAGGCGCTTATATCGAC 3084

QY 987 CAAGG 991
    |||||
Db 3085 CAAGG 3089

```

```

RESULT 8
US-09-974-300-2193
; Sequence 2193, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2193
; LENGTH: 3798
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2193

```

```

Query Match      4.0%; Score 39.8; DB 10; Length 3798;
Best Local Similarity 50.7%; Pred. No. 0.015;
Matches 185; Conservative 0; Mismatches 162; Indels 18; Gaps 3;

QY 630 CTTCCATCGTACATGCCATCCATCGAATCGTTGTTCTGGACAAGATGCCCTCTCAAGCC 689
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2740 CTACCGGCTTACATGGTGGCGCTGCGTATATCATGCTTGAAGAGCTGCCGCTTACCGCT 2799

QY 690 AATGTTAAAGTTGACCGGAAGAACTCTCTCGCAGGGCAAGGTTGTACCGAAGCAGCAG 749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2800 AACGGAAGTCAACCG--ACGCTTCTCCTGAAGCCGATGCCGCCGAATTCACAG 2857

QY 750 ACAGCAGCGCGGTTACCGACATTTCCCATCAGTGAAGTCAATCTTTTCGGAAGAA 809
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2858 AGCGCCGTC-----ACCCGCAATGCTACGGAAGAAAGCTTGCAGTGATT 2904

QY 810 GCCACTGAGGTTGTTGCA---TGAAGTTGACATTTACCGATCACATTTCTTCAATCTCGGT 866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2905 TGGTCTGAGGTTCTCGGCAGACAGCAGATCGGAATAGACGAGAATTTTTTGAATCGG 2964

QY 867 GGACACTCTCTTGGCCACGAAGCTCATTTCTCGTATCGAACACGACTCAAGTCCGT 926
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2965 GGACATTCATTGAAAGCCATGCGCTCGCGGATGCTGAAAGATCTCGCAATCGAT 3024

QY 927 ATCACTGTCAAGGATGCTCTTTGACCATCTGTTTTCGGGATCTAGCATCTGTCTATCG 986
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


RESULT 15
US-10-123-155-536
Sequence 536, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2003, 17:53:27 ; Search time 1708.5 Seconds
(without alignments)
9488.839 Million cell updates/sec

Title: US-09-482-788-1_COPY_7000_8000
Perfect score: 1001
Sequence: 1 aactcgcttttccactcg.....tcgtcaagggtggttg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

| Database : | | | |
|------------|----------------|--|--|
| EST:* | | | |
| 1: | em_estba:* | | |
| 2: | em_esthum:* | | |
| 3: | em_estim:* | | |
| 4: | em_estmu:* | | |
| 5: | em_estov:* | | |
| 6: | em_estpl:* | | |
| 7: | em_estro:* | | |
| 8: | em_hic:* | | |
| 9: | gb_est1:* | | |
| 10: | gb_est2:* | | |
| 11: | gb_hic:* | | |
| 12: | gb_est3:* | | |
| 13: | gb_est4:* | | |
| 14: | gb_est5:* | | |
| 15: | em_estfun:* | | |
| 16: | em_estom:* | | |
| 17: | gb_gss:* | | |
| 18: | em_gss_hum:* | | |
| 19: | em_gss_inv:* | | |
| 20: | em_gss_pln:* | | |
| 21: | em_gss_vrt:* | | |
| 22: | em_gss_fun:* | | |
| 23: | em_gss_mam:* | | |
| 24: | em_gss_mus:* | | |
| 25: | em_gss_other:* | | |
| 26: | em_gss_pro:* | | |
| 27: | em_gss_rod:* | | |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| 1 | 41.2 | 4.1 | 744 | 17 | AQ163140 |
| 2 | 40.4 | 4.0 | 606 | 17 | AQ990481 |
| 3 | 40.4 | 4.0 | 632 | 17 | AQ990701 |
| 4 | 39.8 | 4.0 | 850 | 17 | CNS077YR |
| 5 | 39.8 | 4.0 | 1040 | 17 | CNS078CL |
| 6 | 39.4 | 3.9 | 708 | 17 | AQ989957 |

| | | | | | |
|----|------|-----|------|----|----------|
| 7 | 37.6 | 3.8 | 619 | 10 | BE492765 |
| 8 | 36.6 | 3.7 | 903 | 17 | CNS028CI |
| 9 | 36.4 | 3.6 | 508 | 9 | AL385769 |
| 10 | 36.2 | 3.6 | 342 | 12 | BF809723 |
| 11 | 36.2 | 3.6 | 353 | 12 | BF094070 |
| 12 | 36.2 | 3.6 | 781 | 13 | BM044659 |
| 13 | 36.2 | 3.6 | 790 | 13 | BM048043 |
| 14 | 36.2 | 3.6 | 792 | 13 | BM044732 |
| 15 | 36.2 | 3.6 | 812 | 12 | BG823879 |
| 16 | 36.2 | 3.6 | 828 | 14 | BQ900191 |
| 17 | 36.2 | 3.6 | 901 | 14 | BQ936677 |
| 18 | 36.2 | 3.6 | 949 | 14 | BQ919273 |
| 19 | 36 | 3.6 | 628 | 13 | BI180727 |
| 20 | 35.6 | 3.6 | 602 | 14 | BQ916759 |
| 21 | 35.6 | 3.6 | 812 | 12 | BG319806 |
| 22 | 35.6 | 3.6 | 1050 | 14 | BQ277781 |
| 23 | 35.4 | 3.5 | 653 | 13 | BI174975 |
| 24 | 35 | 3.5 | 360 | 14 | C08494 |
| 25 | 35 | 3.5 | 605 | 10 | BB656627 |
| 26 | 35 | 3.5 | 638 | 10 | AW740666 |
| 27 | 35 | 3.5 | 986 | 17 | CNS00139 |
| 28 | 34.8 | 3.5 | 518 | 9 | AL372468 |
| 29 | 34.8 | 3.5 | 579 | 12 | BE999591 |
| 30 | 34.8 | 3.5 | 579 | 14 | BQ702499 |
| 31 | 34.8 | 3.5 | 749 | 14 | BQ655623 |
| 32 | 34.8 | 3.5 | 782 | 17 | AQ362655 |
| 33 | 34.6 | 3.5 | 330 | 12 | BF745442 |
| 34 | 34.6 | 3.5 | 757 | 9 | AL568342 |
| 35 | 34.4 | 3.4 | 408 | 12 | BF517011 |
| 36 | 34.4 | 3.4 | 437 | 9 | AL377071 |
| 37 | 34.4 | 3.4 | 501 | 14 | C18042 |
| 38 | 34.4 | 3.4 | 942 | 17 | CNS0102V |
| 39 | 34.2 | 3.4 | 333 | 13 | BI192067 |
| 40 | 34.2 | 3.4 | 572 | 9 | AA272108 |
| 41 | 34.2 | 3.4 | 686 | 9 | AL581888 |
| 42 | 34.2 | 3.4 | 792 | 10 | BE052874 |
| 43 | 34.2 | 3.4 | 895 | 12 | BG531334 |
| 44 | 34.2 | 3.4 | 1056 | 17 | CNS022J7 |
| 45 | 34 | 3.4 | 286 | 10 | BB721296 |

ALIGNMENTS

RESULT 1
AQ163140
LOCUS mgxb0023A03r CUGI Rice Blast BAC Library Magnaporthe grisea genomic clone mgxb0023A03r, DNA sequence.
DEFINITION AQ163140
ACCESSION AQ163140.1 GI:3559541
VERSION GSS.
KEYWORDS Magnaporthe grisea.
SOURCE Magnaporthe grisea.
ORGANISM
REFERENCE 1 (bases 1 to 744)
AUTHORS Yu, Y., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R., Phillips, K., Sasinowski, M., Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Magnaporthe grisea Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 431.
Location/Qualifiers

| source | 1. .744 | source | 1. .606 |
|---|---|---|---|
| /organism="Magnaporthe grisea" | /organism="Photobacterium luminescens" | /organism="Photobacterium luminescens" | /organism="Photobacterium luminescens" |
| /strain="70-15" | /strain="W14" | /strain="W14" | /strain="W14" |
| /db_xref="taxon:148305" | /db_xref="taxon:29488" | /db_xref="taxon:29488" | /db_xref="taxon:29488" |
| /clone="mgxb0023A03r" | /clone="PLG01258" | /clone="PLG01258" | /clone="PLG01513" |
| /clone_lib="CUGI Rice Blast BAC Library" | /clone_lib="Photobacterium luminescens strain W14 M13 library" | /clone_lib="Photobacterium luminescens strain W14 M13 library" | /clone_lib="Photobacterium luminescens strain W14 M13 library" |
| /tissue_type="Protoplasts" | /dev_stage="primary phase variant" | /dev_stage="primary phase variant" | /dev_stage="primary phase variant" |
| /lab_host="E. coli DH10B" | /note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus." | /note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus." | /note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus." |
| Rice blast is one of the most devastating fungal diseases of rice world wide. It is a filamentous ascomycete with a haploid genome (n=7) of approximately 40 Mbp. Rice blast is an important model fungal pathogen for studying numerous aspects of the fungal-host interaction. In order to facilitate genome wide analysis, a BAC library containing 9216 clones with an average insert size of 130 kbp was constructed. This library represents greater than 25X genome coverage. High density colony filters are available upon request." | | | |
| BASE COUNT 169 a 243 c 178 g 154 t | BASE COUNT 192 a 122 c 115 g 174 t 3 others | BASE COUNT 189 a 116 c 120 g 204 t 3 others | BASE COUNT 189 a 116 c 120 g 204 t 3 others |
| ORIGIN | ORIGIN | ORIGIN | ORIGIN |
| Query Match | Query Match | Query Match | Query Match |
| Best Local Similarity 4.1%; Score 41.2; DB 17; Length 744; | Best Local Similarity 4.0%; Score 40.4; DB 17; Length 606; | Best Local Similarity 4.0%; Score 40.4; DB 17; Length 606; | Best Local Similarity 4.0%; Score 40.4; DB 17; Length 606; |
| Matches 76; Conservative 0; Mismatches 58; Indels 0; Gaps 0; | Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0; | Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0; | Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0; |
| QY 586 ACCTGTCGATCGCATCGAAGTCCGGAGAGGCTTCGGTCCCTTACTTCATCGTACATGA 645 | QY 835 TTGACATTACCGATCACTTCTTCAATCTCGGTGGACATCTCTCTTGGCCAGCAAGCTCA 894 | QY 835 TTGACATTACCGATCACTTCTTCAATCTCGGTGGACATCTCTCTTGGCCAGCAAGCTCA 894 | QY 835 TTGACATTACCGATCACTTCTTCAATCTCGGTGGACATCTCTCTTGGCCAGCAAGCTCA 894 |
| DB 71 ATCGTGTATGAGGCTGAGCTTCGCGACCATCTCAAGTTCGCGACCTAGCGCG 130 | DB 268 TTGGATTATATGATAATTTCTTCTCACTCGGTGGTCACTCTTAGCAGTTAAATAA 327 | DB 268 TTGGATTATATGATAATTTCTTCTCACTCGGTGGTCACTCTTAGCAGTTAAATAA 327 | DB 268 TTGGATTATATGATAATTTCTTCTCACTCGGTGGTCACTCTTAGCAGTTAAATAA 327 |
| QY 646 TCCATCGAATCGTTGTTCTGGACAAGATGCTCTCAAGCCAAATGTTAAAGTTGACC 705 | QY 895 TTTCTCGTATCGACCAACGACTCAAGTCCGTATCACTGTCGAAGATGCTTTTGACCATC 954 | QY 895 TTTCTCGTATCGACCAACGACTCAAGTCCGTATCACTGTCGAAGATGCTTTTGACCATC 954 | QY 895 TTTCTCGTATCGACCAACGACTCAAGTCCGTATCACTGTCGAAGATGCTTTTGACCATC 954 |
| DB 131 TTCTAGCATCTTCATTTCTGCTGAGAGTGCTTAATTAACCCCAACGCGAAGTTGACA 190 | DB 328 CGCTGCTATAGAATAATGTTAAACAATCATTTAGTATGATGATGTTGTTGGCAATC 387 | DB 328 CGCTGCTATAGAATAATGTTAAACAATCATTTAGTATGATGATGTTGTTGGCAATC 387 | DB 328 CGCTGCTATAGAATAATGTTAAACAATCATTTAGTATGATGATGTTGTTGGCAATC 387 |
| QY 706 GGAAGGAACCTCTCT 719 | QY 955 CTGTATTTTCGGGATCTAGCATCTGTCATCCGTC 988 | QY 955 CTGTATTTTCGGGATCTAGCATCTGTCATCCGTC 988 | QY 955 CTGTATTTTCGGGATCTAGCATCTGTCATCCGTC 988 |
| DB 191 AGCCCAATCTACT 204 | DB 388 CAACCATGTCCTAGATTCGCGATTCGATTCGCGCA 421 | DB 388 CAACCATGTCCTAGATTCGCGATTCGATTCGCGCA 421 | DB 388 CAACCATGTCCTAGATTCGCGATTCGATTCGCGCA 421 |
| RESULT 2 | RESULT 3 | RESULT 3 | RESULT 3 |
| AQ990481 | AQ990701/c | AQ990701/c | AQ990701/c |
| LOCUS | LOCUS | LOCUS | LOCUS |
| DEFINITION | DEFINITION | DEFINITION | DEFINITION |
| ACCESSION | ACCESSION | ACCESSION | ACCESSION |
| VERSION | VERSION | VERSION | VERSION |
| KEYWORDS | KEYWORDS | KEYWORDS | KEYWORDS |
| SOURCE | SOURCE | SOURCE | SOURCE |
| ORGANISM | ORGANISM | ORGANISM | ORGANISM |
| REFERENCE | REFERENCE | REFERENCE | REFERENCE |
| AUTHORS | AUTHORS | AUTHORS | AUTHORS |
| TITLE | TITLE | TITLE | TITLE |
| JOURNAL | JOURNAL | JOURNAL | JOURNAL |
| MEDLINE | MEDLINE | MEDLINE | MEDLINE |
| COMMENT | COMMENT | COMMENT | COMMENT |
| 1 (bases 1 to 606) | 1 (bases 1 to 632) | 1 (bases 1 to 632) | 1 (bases 1 to 632) |
| french-Constant R.H., Waterfield,N., Burland,V., Perna,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R. | french-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R. | french-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R. | french-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R. |
| A genomic sample sequence of the entomopathogenic bacterium Photobacterium luminescens W14: potential implications for virulence | A genomic sample sequence of the entomopathogenic bacterium Photobacterium luminescens W14: potential implications for virulence | A genomic sample sequence of the entomopathogenic bacterium Photobacterium luminescens W14: potential implications for virulence | A genomic sample sequence of the entomopathogenic bacterium Photobacterium luminescens W14: potential implications for virulence |
| Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000) | Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000) | Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000) | Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000) |
| Contact: french-Constant RH | Contact: french-Constant RH | Contact: french-Constant RH | Contact: french-Constant RH |
| Department of Biology and Biochemistry | Department of Biology and Biochemistry | Department of Biology and Biochemistry | Department of Biology and Biochemistry |
| University of Bath | University of Bath | University of Bath | University of Bath |
| South Building, Bath BA2 7AY, UK | South Building, Bath BA2 7AY, UK | South Building, Bath BA2 7AY, UK | South Building, Bath BA2 7AY, UK |
| Tel: (44) 1225 826621 | Tel: (44) 1225 826621 | Tel: (44) 1225 826621 | Tel: (44) 1225 826621 |
| Fax: (44) 1225 826779 | Fax: (44) 1225 826779 | Fax: (44) 1225 826779 | Fax: (44) 1225 826779 |
| Email: bssrfc@bath.ac.uk | Email: bssrfc@bath.ac.uk | Email: bssrfc@bath.ac.uk | Email: bssrfc@bath.ac.uk |
| This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res. | This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res. | This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res. | This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res. |
| Seq primer: M13 Forward | Seq primer: M13 Forward | Seq primer: M13 Forward | Seq primer: M13 Forward |
| Class: shotgun. | Class: shotgun. | Class: shotgun. | Class: shotgun. |
| Location/Qualifiers | Location/Qualifiers | Location/Qualifiers | Location/Qualifiers |
| 1. .606 | 1. .632 | 1. .632</ | |

5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
1..1040
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="BB0AA019F07"
/clone_lib="BB0AA"
/notes="end : T7"
misc_feature
8..>1027
/notes="similar to Saccharomyces cerevisiae ORF YBR115c [LY52 : L-aminoadipate-semialdehyde dehydrogenase, large subunit 1]"
/evidence="not experimental"
BASE COUNT 256 a 262 c 287 g 234 t 1 others
ORIGIN

Query Match 4.0%; Score 39.8; DB 17; Length 1040;
Best Local Similarity 56.5%; Pred. No. 3.3;
Matches 74; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 852 TTCTTCAATCTCGTGGACACTCTCTTGGCCACGAGCTCATTTCTCTATCGACCAA 911
||||| ||| ||||||||||| ||| ||||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 602 TTCTTGACCTGGTGGACACTCGATCCTCGCCACGCGGATGATTTTGGATTGAGAA 661
QY 912 CGACTCAAGTCGCTCACTGTCAAGATGCTTTGACCATCTGTTATTTGGGATCTA 971
||||| ||| ||||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| ||||| |||
Db 662 CAGCTTCCCTCATATCTCTCTCGGAGCCATCTTCAACACCCACCGCTTGTTCATTC 721
QY 972 GCATCTGTCTAT 982
||||| ||| ||| |||
Db 722 GCAGCAGAGAT 732

RESULT 6
AQ989957
LOCUS 708 bp DNA linear GSS 14-AUG-2000
DEFINITION R1C00641 Photorhabdus luminescens strain W14 M13 library
Photorhabdus luminescens genomic clone PLG00641, DNA sequence.
ACCESSION AQ989957
VERSION AQ989957.1 GI:9648551
KEYWORDS GSS.
SOURCE Photorhabdus luminescens.
ORGANISM Photorhabdus luminescens.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Photorhabdus.
REFERENCE 1 (bases 1 to 708)
AUTHORS ffrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens w14: potential implications for virulence
JOURNAL Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE 20378633
COMMENT Contact: ffrench-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bssr@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic Acids Res.
Seq primer: M13 Forward
Class: Shotgun.
Location/Qualifiers
1..708
/organism="Photorhabdus luminescens"
/strain="W14"
/db_xref="taxon:29488"

FEATURES

source
1..708
/organism="Photorhabdus luminescens"
/strain="W14"
/db_xref="taxon:29488"

/clone="PLG00641"
/clone_lib="Photorhabdus luminescens strain W14 M13 library"
/dev_stage="primary phase variant"
/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."
BASE COUNT 220 a 150 c 152 g 183 t 3 others
ORIGIN

Query Match 3.9%; Score 39.4; DB 17; Length 708;
Best Local Similarity 52.8%; Pred. No. 3.1;
Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 814 CTGAGGTGTTGGCATGAAGTTTCACATTACCGATCATCTTCTCAATCTCGGTGACACT 873
||||| ||| ||||||||||| ||| ||||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 18 CAGTGTGGCGGTAGTCAAGTCGGCATTTATGACAACTTTTCGTCCTGGTGTCACT 77
QY 874 CTCTCTTGGCCACGAGCTCATTTCTGTCGTCGACCAACGACTCAAGGTCGGTATCACTG 933
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 CTCTTCAAGCAACCGCGCTTATCAGCTTAATACGCAATGATCTAAACATTGAAATACCGT 137
QY 934 TCAGGATGCTTTGACCATCTCTATTTCGGGATCTAGCA 974
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 TGACGACTGCTTTGAACATCCAACTTGAGCGATTAGCA 178

RESULT 7
BE492765 619 bp mRNA linear EST 16-APR-2001
LOCUS WHE0564_E04_E04ZE Triticum monococcum vegetative apex cDNA library
DEFINITION Triticum monococcum cDNA clone WHE0564_E04_E04, mRNA sequence.
ACCESSION BE492765
VERSION BE492765.1 GI:9659358
KEYWORDS EST.
SOURCE Triticum monococcum.
ORGANISM Triticum monococcum.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 619)
AUTHORS Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Stamov,B. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - Vegetative apex cDNA library from Triticum monococcum Unpublished (2001)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: candersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene T3 primer.
Location/Qualifiers
1..619
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/clone="WHE0564_E04_E04"
/clone_lib="Triticum monococcum vegetative apex cDNA library"
/tissue_type="Vegetative shoot apex"
/dev_stage="Three weeks-old plants"
/lab_host="E. coli XL0LR"
/note="vector: Lambda pBK-CMV (Lambda Zap Express), excised phagemid; Site_1: EcoRI; Site_2: XhoI; The tissue total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised at the University of California, Davis (V. Echenique, B. Stamova, J. Dubcovsky). Plasmid DNA preparations and DNA

FEATURES
source
1..619
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/clone="WHE0564_E04_E04"
/clone_lib="Triticum monococcum vegetative apex cDNA library"
/tissue_type="Vegetative shoot apex"
/dev_stage="Three weeks-old plants"
/lab_host="E. coli XL0LR"
/note="vector: Lambda pBK-CMV (Lambda Zap Express), excised phagemid; Site_1: EcoRI; Site_2: XhoI; The tissue total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised at the University of California, Davis (V. Echenique, B. Stamova, J. Dubcovsky). Plasmid DNA preparations and DNA

ORIGIN


```

Query Match          3.6%; Score 36.2; DB 12; Length 353;
Best Local Similarity 53.1%; Pred. No. 15;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 798 CTTTCCGAAGCCACTGAGGTGTTGGCATGAAGTTGACATACCGATCACTTCTTC 857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CTGGTCTATGATTTCTCGGTGTAGACAACTCCAAAGTGGCTTTGCCAATGACTCGGTC 122

QY 858 AATCTCGGTGGACACTCTCTCTTGGCCACGAGCTCATTTCTGATCGACCAAGCACTC 917
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ACTGTGGTGACCAACCGCTCATGGGAATCAGCACATCCTTGACCTCAGCCAGAGCGC 62

QY 918 AAGTCCGATATCACTGTCAAGGATG 942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 AGAGTCCAGGTCCTTAGCTGGATG 37

RESULT 12
BM044659/c
LOCUS
DEFINITION 603622365F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5447729 5',
            mRNA sequence.
ACCESSION BM044659
VERSION   BM044659.1 GI:16773926
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 781)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: DCTD/DRP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1932 row: b column: 18
            High quality sequence start: 4
            High quality sequence stop: 762.
FEATURES
    Location/Qualifiers
        1..781
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5447729"
            /clone_lib="NIH_MGC_40"
            /tissue_type="carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
            Site:2: EcoRI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Library constructed by
            Ling Hong in the laboratory of Gerald M. Rubin (University
            of California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC Library."
BASE COUNT 168 a 214 c 223 g 176 t
ORIGIN

Query Match          3.6%; Score 36.2; DB 13; Length 781;
Best Local Similarity 53.1%; Pred. No. 31;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 798 CTTTCCGAAGCCACTGAGGTGTTGGCATGAAGTTGACATACCGATCACTTCTTC 857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 CTGGTCTATGATTTCTCGGTGTAGACAACTCCAAAGTGGCTTTGCCAATGACTCGGTC 140

RESULT 14
Query Match          3.6%; Score 36.2; DB 13; Length 790;
Best Local Similarity 53.1%; Pred. No. 31;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 798 CTTTCCGAAGCCACTGAGGTGTTGGCATGAAGTTGACATACCGATCACTTCTTC 857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 CTGGTCTATGATTTCTCGGTGTAGACAACTCCAAAGTGGCTTTGCCAATGACTCGGTC 141

QY 858 AATCTCGGTGGACACTCTCTTGGCCACGAGCTCATTTCTGATCGACCAAGCACTC 917
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 ACTGTGGTGACCAACCGCTCATGGGAATCAGCACATCCTTGACCTCAGCCAGAGCGC 81

QY 918 AAGTCCGATATCACTGTCAAGGATG 942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 AGAGTCCAGGTCCTTAGCTGGATG 56

RESULT 14

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:33:15 ; Search time 115 Seconds
(without alignments)
3625.576 Million cell updates/sec

Title: US-09-482-788-2

Perfect score: 16128

Sequence: 1 MEYLFVAVDGRDLPTTASF.....RVEHLLVEVSKTFEGLNSSL 3129

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-------------|--------------------|
| 1 | 16128 | 100.0 | 3129 | 21 AAB07427 | Amino acid sequenc |
| 2 | 8799 | 54.6 | 3210 | 22 AAB73958 | Mycelia sterilia c |
| 3 | 5000 | 31.0 | 15281 | 15 AAR44929 | T. niveum Cyclospo |
| 4 | 2146.5 | 13.3 | 2841 | 21 AAB07581 | Protein encoded by |
| 5 | 1785.5 | 11.1 | 3587 | 14 AAR34713 | Bacillus subtilis |
| 6 | 1713.5 | 10.6 | 2448 | 22 AAU36277 | Pseudomonas aerugi |
| 7 | 1677 | 10.4 | 4342 | 22 AAU33611 | Pseudomonas aerugi |
| 8 | 1632 | 10.1 | 4999 | 23 AAO22159 | Ramoplanin biosynt |
| 9 | 1587.5 | 9.8 | 3588 | 14 AAR34712 | Bacillus subtilis |
| 10 | 1572 | 9.7 | 1537 | 22 AAB83971 | Amino acid sequenc |

| | | | | | |
|----|--------|-----|------|-------------|--------------------|
| 11 | 1516 | 9.4 | 3639 | 14 AAR40227 | ACVS. Acremonium |
| 12 | 1515 | 9.4 | 3712 | 12 AAR13896 | ACV synthetase. A |
| 13 | 1429 | 8.9 | 2323 | 22 AAG81833 | S. epidermidis ope |
| 14 | 1425.5 | 8.8 | 2404 | 23 ABP38619 | Staphylococcus epi |
| 15 | 1413.5 | 8.8 | 3722 | 12 AAR10145 | Cephalosporin anti |
| 16 | 1406 | 8.7 | 2397 | 22 AAU36672 | Staphylococcus aur |
| 17 | 1375.5 | 8.5 | 4999 | 23 AAO22158 | Ramoplanin biosynt |
| 18 | 1374 | 8.5 | 1668 | 22 AAU34158 | Staphylococcus aur |
| 19 | 1335 | 8.3 | 3778 | 12 AAR13895 | ACV synthetase. P |
| 20 | 1332.5 | 8.3 | 3768 | 12 AAR13753 | ACVS. Penicillium |
| 21 | 1298 | 8.0 | 2162 | 21 AAB07560 | Protein encoded by |
| 22 | 1048.5 | 6.5 | 1591 | 22 ABB52772 | Escherichia coli p |
| 23 | 1006.5 | 6.2 | 2766 | 22 AAB83972 | Amino acid sequenc |
| 24 | 968 | 6.0 | 2675 | 21 AAB07564 | Protein encoded by |
| 25 | 914.5 | 5.7 | 1218 | 21 AAB07563 | Protein encoded by |
| 26 | 876 | 5.4 | 2626 | 21 AAB07569 | Protein encoded by |
| 27 | 791 | 4.9 | 1066 | 21 AAB07561 | Protein encoded by |
| 28 | 782.5 | 4.9 | 1410 | 21 AAY58574 | Sorangium cellulos |
| 29 | 759 | 4.7 | 2392 | 21 AAB07565 | Amino acid sequenc |
| 30 | 756 | 4.7 | 1274 | 14 AAR34714 | Bacillus subtilis |
| 31 | 684.5 | 4.2 | 766 | 22 AAG81663 | S. epidermidis ope |
| 32 | 675 | 4.2 | 1874 | 22 ABB52779 | Escherichia coli p |
| 33 | 641 | 4.0 | 1435 | 22 ABB52769 | Escherichia coli p |
| 34 | 623 | 3.9 | 1051 | 23 AAO22157 | Ramoplanin biosynt |
| 35 | 582.5 | 3.6 | 760 | 22 ABB52770 | Escherichia coli p |
| 36 | 571.5 | 3.5 | 1103 | 22 AAU52523 | Propionibacterium |
| 37 | 559 | 3.5 | 1293 | 22 AAU34486 | E. coli cellular p |
| 38 | 554 | 3.4 | 1455 | 22 ABB52766 | Escherichia coli p |
| 39 | 538.5 | 3.3 | 1294 | 22 AAU38131 | Salmonella typhi c |
| 40 | 507 | 3.1 | 450 | 23 AAE14387 | Human AMP-binding |
| 41 | 490 | 3.0 | 578 | 21 AAB07587 | Protein encoded by |
| 42 | 475.5 | 2.9 | 891 | 23 AAO22162 | Ramoplanin biosynt |
| 43 | 434.5 | 2.7 | 1391 | 21 AAY44644 | Candida albicans a |
| 44 | 424 | 2.6 | 1274 | 22 AAB79202 | Corynebacterium ql |
| 45 | 424 | 2.6 | 1295 | 22 AAG92726 | C glutamicum prote |

ALIGNMENTS

RESULT 1

AAB07427

ID AAB07427 standard; Protein; 3129 AA.

XX AAB07427;

XX AC

XX 20-OCT-2000 (first entry)

XX DE Amino acid sequence of a cyclohexadepsipeptide synthetase.

XX KW Cyclohexadepsipeptide synthetase; filamentous fungal cell;

XX KW cyclohexadepsipeptide; antibiotic.

XX OS Fusarium venenatum.

XX XX

XX WO200042203-A2.

XX PD 20-JUL-2000.

XX XX

XX 13-JAN-2000; 2000WO-US00913.

XX XX

XX 13-JAN-1999; 99US-0229862.

XX XX

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX PA

XX PI Berka RM, Rey MW, Yoder WT;

XX XX

XX DR WPI; 2000-482833/42.

XX DR N-PSDB; AAA58762.

XX XX

XX PT Producing a heterologous polypeptide for production of antibiotics

XX PT comprises cultivating a mutant of a parent filamentous fungal cell

XX PT comprising a nucleic acid sequence encoding cyclohexadepsipeptide -

xx
ps
xx

Claim 30; Fig 1A-1; 76pp; English.

CC The present sequence represents a cyclohexadepsipeptide synthetase polypeptide. The specification describes a method for producing a heterologous polypeptide. The method comprises cultivating a mutant of a parent filamentous fungal cell, which produces less cyclohexadepsipeptide than the parent filamentous fungal cell when cultured under the same conditions. The method if used for the production of biologically active compounds e.g. antibiotics.

xx
sq

Sequence 3129 AA;

Query Match 100.0%; Score 16128; DB 21; Length 3129;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEYLTAVDGRODLPTTPASFCSHGDSPLNSSEYOLFHLGLDSSRIEAIKPTCFQDLM 60
 Db 1 MEYLTAVDGRODLPTTPASFCSHGDSPLNSSEYOLFHLGLDSSRIEAIKPTCFQDLM 60

Qy 61 DCNALDKOSAIGHAVYDPTDIDISRFALAWKEIVNQTALRAFAFTSDSGKTSQVILKD 120
 Db 61 DCNALDKOSAIGHAVYDPTDIDISRFALAWKEIVNQTALRAFAFTSDSGKTSQVILKD 120

Qy 121 SFVSWCWSSSSPDEWVDEAAAAAGPRCNRFVLLDMQTKKCOLVWTFSHALVDVT 180
 Db 121 SFVSWCWSSSSPDEWVDEAAAAAGPRCNRFVLLDMQTKKCOLVWTFSHALVDVT 180

Qy 181 FQORVLSRVFAAYKHEKDHTRPETPSSDATDTSQSVSVSMSCEDNAVSAATHFWQTHL 240
 Db 181 FQORVLSRVFAAYKHEKDHTRPETPSSDATDTSQSVSVSMSCEDNAVSAATHFWQTHL 240

Qy 241 NDLNASVPHLSHLVNPNTTAEHRITPPLSKOKALSNAICRTALSILLSRYTHSDEA 300
 Db 241 NDLNASVPHLSHLVNPNTTAEHRITPPLSKOKALSNAICRTALSILLSRYTHSDEA 300

Qy 301 LFGAVTEOSLPFDKHYLDGTYQTVAPLRVHCQSNLRASDVMDAISYDDRLGLHAPFL 360
 Db 301 LFGAVTEOSLPFDKHYLDGTYQTVAPLRVHCQSNLRASDVMDAISYDDRLGLHAPFL 360

Qy 361 RDIRNTGDSGAACDFOTVLLVTDGSHVNNINGFLOQITESSHFMCNNRALLHCOME 420
 Db 361 RDIRNTGDSGAACDFOTVLLVTDGSHVNNINGFLOQITESSHFMCNNRALLHCOME 420

Qy 421 SSGALLVAYYDHNVIDSLQTRLLQORGHILKCLQSPDLSSMAEVNLMTEYDRAETESW 480
 Db 421 SSGALLVAYYDHNVIDSLQTRLLQORGHILKCLQSPDLSSMAEVNLMTEYDRAETESW 480

Qy 481 NSQPLEVQDTLIHHEMLKAVSHSPTKTAIOAGDGDWYSELNDVSSRLAVHIKSLGLRAQ 540
 Db 481 NSQPLEVQDTLIHHEMLKAVSHSPTKTAIOAGDGDWYSELNDVSSRLAVHIKSLGLRAQ 540

Qy 541 QAIIPVYFEKSKWVIASMLAVLKSNAFTLIDNPDPARTAQVVTQTRATVALTSKLHRE 600
 Db 541 QAIIPVYFEKSKWVIASMLAVLKSNAFTLIDNPDPARTAQVVTQTRATVALTSKLHRE 600

Qy 601 TVQKLVGRVYVDELLOQSVASDDSSLTQKSDLAYVFTSGSTGPKGIMIEHRAFS 660
 Db 601 TVQKLVGRVYVDELLOQSVASDDSSLTQKSDLAYVFTSGSTGPKGIMIEHRAFS 660

Qy 661 CALKFGASLGINSDFALQFGTHAFAGACILLEIMTTLINGGCVCIPSDDDRMNSIPSPINR 720
 Db 661 CALKFGASLGINSDFALQFGTHAFAGACILLEIMTTLINGGCVCIPSDDDRMNSIPSPINR 720

Qy 721 YNNWNMMATPSYMGTFSPEDVPGLATLVLVGEOMSSSVNAIWPAPKQLQNGYQSESSSI 780
 Db 721 YNNWNMMATPSYMGTFSPEDVPGLATLVLVGEOMSSSVNAIWPAPKQLQNGYQSESSSI 780

Qy 781 CFASNNSTENNNCRAGVCAHSWIDPNDINRLVPIGAVGELVTVESPGIARDYIVPPPEK 840
 Db 781 CFASNNSTENNNCRAGVCAHSWIDPNDINRLVPIGAVGELVTVESPGIARDYIVPPPEK 840

Qy 841 SPFFTDIPSWYPANTFPDGAKYRTGDLARVASDGSIVCLGRIDSQVKIRQORVELGAIE 900
 Db 841 SPFFTDIPSWYPANTFPDGAKYRTGDLARVASDGSIVCLGRIDSQVKIRQORVELGAIE 900

Qy 901 THLRQOQWDDLTIVVEATKRSQSANSTSLIAFLTGSSYFGNRPDSADHLHDHDAKAINIK 960
 Db 901 THLRQOQWDDLTIVVEATKRSQSANSTSLIAFLTGSSYFGNRPDSADHLHDHDAKAINIK 960

Qy 961 LEQVLPKHSIPSFYICMLELPTATGKIDRRRLRMKGDIIDKOTQGAIVQOAPAPIPVF 1020
 Db 961 LEQVLPKHSIPSFYICMLELPTATGKIDRRRLRMKGDIIDKOTQGAIVQOAPAPIPVF 1020

Qy 1021 ADTAAKLHSIWQSLIGIDPATVNVGATFFELGGNSITAIRMVNMAVSVMGLKVSNIYQH 1080
 Db 1021 ADTAAKLHSIWQSLIGIDPATVNVGATFFELGGNSITAIRMVNMAVSVMGLKVSNIYQH 1080

Qy 1081 PTLAGISAVKVGDPPLSYTLIPKSTHEGPVEQSYSGRLWFLDQDLVGSWLWLIPIYAVRMR 1140
 Db 1081 PTLAGISAVKVGDPPLSYTLIPKSTHEGPVEQSYSGRLWFLDQDLVGSWLWLIPIYAVRMR 1140

Qy 1141 GPNVVDALRRALAALEQRHETLRTTFEDQCGVGQIVHEKLSSEMKVIDLCSGLDPEV 1200
 Db 1141 GPNVVDALRRALAALEQRHETLRTTFEDQCGVGQIVHEKLSSEMKVIDLCSGLDPEV 1200

Qy 1201 LNQEQTTPFNLSSEAGWRATLLRLGEDDHILTIYVHHIISDGWSIDVLRRDLNOLYSAAL 1260
 Db 1201 LNQEQTTPFNLSSEAGWRATLLRLGEDDHILTIYVHHIISDGWSIDVLRRDLNOLYSAAL 1260

Qy 1261 KSKDPLSALTPLPIQYSDFAKWQDOFIEQKOLNYKKOLKQSSPAKIPDFAFPALL 1320
 Db 1261 KSKDPLSALTPLPIQYSDFAKWQDOFIEQKOLNYKKOLKQSSPAKIPDFAFPALL 1320

Qy 1321 SGDAGCVHVITIDGELYQSLRAFPCNEHNTTSFVLLAAAFRAAHYRLTAVEDAVICTPIANR 1380
 Db 1321 SGDAGCVHVITIDGELYQSLRAFPCNEHNTTSFVLLAAAFRAAHYRLTAVEDAVICTPIANR 1380

Qy 1381 NRPELEDIICGFVNTQCMRINIDHDDTFTGLINOVKATTTAAFNEDIPFVRVVSALQPG 1440
 Db 1381 NRPELEDIICGFVNTQCMRINIDHDDTFTGLINOVKATTTAAFNEDIPFVRVVSALQPG 1440

Qy 1441 SRDLSSTPLAQLIFAVHSQKDLGRFKFQGLSVPSKAYTRFDMFHLFQETDSLKGSV 1500
 Db 1441 SRDLSSTPLAQLIFAVHSQKDLGRFKFQGLSVPSKAYTRFDMFHLFQETDSLKGSV 1500

Qy 1501 NFADELFKMETVENVRVFFELIRNGLQSSRTSPVSIPLTDGIVTLBKLDVLNVKHVDYP 1560
 Db 1501 NFADELFKMETVENVRVFFELIRNGLQSSRTSPVSIPLTDGIVTLBKLDVLNVKHVDYP 1560

Qy 1561 RESSLADVFOTQVSAYPDSLAVVDSSCRLTYTELDRQSDILAGLWLRSSMPAETLVAVFA 1620
 Db 1561 RESSLADVFOTQVSAYPDSLAVVDSSCRLTYTELDRQSDILAGLWLRSSMPAETLVAVFA 1620

Qy 1621 PRSCETIVAFGVLKANLAYLPLDVRSPSARVQDILSGLSPTIVLIGHDTAPDIEVTN 1680
 Db 1621 PRSCETIVAFGVLKANLAYLPLDVRSPSARVQDILSGLSPTIVLIGHDTAPDIEVTN 1680

Qy 1681 VEFVRIDALNDSNADGFVEIHDSTKPSATSLAYVLYTSGSTGRPKGVMIHVRVIRTV 1740
 Db 1681 VEFVRIDALNDSNADGFVEIHDSTKPSATSLAYVLYTSGSTGRPKGVMIHVRVIRTV 1740

Qy 1741 TSCCIPNYPSETMAHMTATFAGDASVEIYISALLFGRTLVCVDYMTTLDARALKDVPFRE 1800
 Db 1741 TSCCIPNYPSETMAHMTATFAGDASVEIYISALLFGRTLVCVDYMTTLDARALKDVPFRE 1800

Qy 1801 HVNAAHSVTSQDVPRLVRPRLSRTLMFFFLVVVTDSTAPDALDAQGLYGVQCYNGYGP 1860
 Db 1801 HVNAAHSVTSQDVPRLVRPRLSRTLMFFFLVVVTDSTAPDALDAQGLYGVQCYNGYGP 1860

Qy 1861 TENGVMSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEOQLVGIGVMGELVVTGDGLAR 1920
 Db 1861 TENGVMSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEOQLVGIGVMGELVVTGDGLAR 1920

Qy 1921 GYSDKALDENRFRVHITVNDQTVKAYRTGDRVRYRIGDGLIEFFGRMDTQFKIRGNRTESA 1980

```

1921 GYSKALDENRFVHTVNDQTVKAYRTGDRYVRIGDGLIEFFGRMDTQFKRGNRIESA 1980
1981 EIEAALLRDSVRDAAVVLOQNEOAPILGFVVADHDHSDKQSANOVGQMDHFE 2040
1981 EIEAALLRDSVRDAAVVLOQNEOAPILGFVVADHDHSDKQSANOVGQMDHFE 2040
2041 GMYSDIGIDPSTIGSDFKGWTSMYDGSQIDFDEMHEWLGETTRTLHDNRSLGNVLEIGT 2100
2041 GMYSDIGIDPSTIGSDFKGWTSMYDGSQIDFDEMHEWLGETTRTLHDNRSLGNVLEIGT 2100
2101 GSGMILFNLDRLSLEYGLERSRNSAAAFVKNATESIPSLAGKAKVQVGTATDICOVDLH 2160
2101 GSGMILFNLDRLSLEYGLERSRNSAAAFVKNATESIPSLAGKAKVQVGTATDICOVDLH 2160
2161 PDLVVLNSVIOYFSSSEYLAELADTLIHLPNVQRIFFGDRVSQATNEHFLAARAIHTLCK 2220
2161 PDLVVLNSVIOYFSSSEYLAELADTLIHLPNVQRIFFGDRVSQATNEHFLAARAIHTLCK 2220
2221 NATKDDVRQKMAELEDMEELLVEPAFTSLKDRPGLVHVEILPKNMEAVNELSAYRY 2280
2221 NATKDDVRQKMAELEDMEELLVEPAFTSLKDRPGLVHVEILPKNMEAVNELSAYRY 2280
2281 AAVHVHRSGLGDELVLPEKDDWIDFOANQLNOKSLGDLKSSDAAINAVSKIPPEITAF 2340
2281 AAVHVHRSGLGDELVLPEKDDWIDFOANQLNOKSLGDLKSSDAAINAVSKIPPEITAF 2340
2341 ERQVVASLNSNIDEMQSLTIRSSAEGSSLSVPDIFRIAGAGFRVEYSSAROWSONGAL 2400
2341 ERQVVASLNSNIDEMQSLTIRSSAEGSSLSVPDIFRIAGAGFRVEYSSAROWSONGAL 2400
2401 DAVFHCCSQGRTLVNFTDHLRGSLTLNRQLQNRRIAIEVRERLSLPSYMP 2460
2401 DAVFHCCSQGRTLVNFTDHLRGSLTLNRQLQNRRIAIEVRERLSLPSYMP 2460
2461 SNIVLDKMPNLNANGKVDKELSRRAKVVVQKQTAAPLTPPISEVEVILCEEATEVFGM 2520
2461 SNIVLDKMPNLNANGKVDKELSRRAKVVVQKQTAAPLTPPISEVEVILCEEATEVFGM 2520
2521 KVDITDFHFNGLGSHSLATKILSRIDQRLKVRITVKDVFDPVADLASVIRQGLGLOOP 2580
2521 KVDITDFHFNGLGSHSLATKILSRIDQRLKVRITVKDVFDPVADLASVIRQGLGLOOP 2580
2581 VSDGQGDRSAHMAPRTETAILCDEFKVLGQVITDNFDFLGGHSLMATKLAIRIGH 2640
2581 VSDGQGDRSAHMAPRTETAILCDEFKVLGQVITDNFDFLGGHSLMATKLAIRIGH 2640
2641 RLDTTVSKVDVDFHPVLFQALALDNLVQSKTNEIVGGREMAEYSPFQLLTEDPEEFMA 2700
2641 RLDTTVSKVDVDFHPVLFQALALDNLVQSKTNEIVGGREMAEYSPFQLLTEDPEEFMA 2700
2701 SEIKPOLBELQETIIDIYSTOMQKAFLEDHTTARPPVPVFFIDFPSTSEPDAAGLIRAC 2760
2701 SEIKPOLBELQETIIDIYSTOMQKAFLEDHTTARPPVPVFFIDFPSTSEPDAAGLIRAC 2760
2761 ESLVNLHDFIRTVFAEASGELYQVVLSCDLPLQVETEDNTINTATNEFLDEFAKEPVRL 2820
2761 ESLVNLHDFIRTVFAEASGELYQVVLSCDLPLQVETEDNTINTATNEFLDEFAKEPVRL 2820
2821 GHPLIRFTIINKTSMRVMIRISHALYDGLSLEHVVRKLHMLYNGRSLLPHPQFSRYQY 2880
2821 GHPLIRFTIINKTSMRVMIRISHALYDGLSLEHVVRKLHMLYNGRSLLPHPQFSRYQY 2880
2881 TADGRESHGFWRDVQIOTNPTILSDDTVWVGNDATCKALHLSKLVNIPSOVLGRSSNII 2940
2881 TADGRESHGFWRDVQIOTNPTILSDDTVWVGNDATCKALHLSKLVNIPSOVLGRSSNII 2940
2941 TOATVFNACALVLSRESKSDVWFGRTVSRQGLPVEYQDVGCTNAVPVRAHIESSD 3000
2941 TOATVFNACALVLSRESKSDVWFGRTVSRQGLPVEYQDVGCTNAVPVRAHIESSD 3000
3001 YNQLLHDIDQOYLISLPHETIGFSDLKRNCTDWPDAITNFSCCITYHNFEYHPESQFEQ 3060
3001 YNQLLHDIDQOYLISLPHETIGFSDLKRNCTDWPDAITNFSCCITYHNFEYHPESQFEQ 3060

Db 3001 YNQLLHDIDQOYLISLPHETIGFSDLKRNCTDWPDAITNFSCCITYHNFEYHPESQFEQ 3060
Qy 3061 RVEMGVLTAKFVNIEMDEPLDIAIAGEVEPDGAGLKVTVIKTQLFGKRKRVEHLLLEVSK 3120
Db 3061 RVEMGVLTAKFVNIEMDEPLDIAIAGEVEPDGAGLKVTVIKTQLFGKRKRVEHLLLEVSK 3120
Qy 3121 TFEGLNSSL 3129
Db 3121 TFEGLNSSL 3129

RESULT 2
AAB73958
ID AAB73958 standard; Protein; 3210 AA.
XX
AC AAB73958;
XX
DT 29-MAY-2001 (first entry)
XX
Mycelia sterilia cyclic depsipeptide synthase.
XX
Mycelia sterilia; cyclic depsipeptide synthase; anthelmintic;
KW PF1022.
XX
Mycelia sterilia.
OS
XX WO200118179-A1.
PN
XX 15-MAR-2001.
PD
XX
PF 07-SEP-2000; 2000WO-JP06103.
XX
PR 07-SEP-1999; 99JP-0253040.
PR 06-APR-2000; 2000JP-0104291.
XX
PA (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX
PI Midoh N, Okakura K, Miyamoto K, Watanabe M, Yanai K, Yasutake T;
Aihara S, Futamura T, Kleinkauf H, Murakami T;
XX
DR WPI: 2001-265970/27.
DR N-PSDB; AAF79702.
XX
Novel cyclic depsipeptide synthase and gene encoding it for efficient
production of anthelmintic substance PF1022
XX
Claim 1; Page 59-80; 92pp; Japanese.
XX
The present sequence is a cyclic depsipeptide synthase from Mycelia
sterilia. The present protein, or a protein containing an addition,
deletion and/or substitution of one or more amino acid residues is useful
for the efficient production of the anthelmintic PF1022 (cyclo(D-lactyl-
L-N-methylleucyl-D-3-phenyllactyl-L-N-methylleucyl-D-lactyl-L-N-
methylleucyl-D-3-phenyllactyl-L-N-methylleucyl)).
XX
Sequence 3210 AA;
Query Match 54.6%; Score 8799; DB 22; Length 3210;
Best Local Similarity 54.6%; Pred. No. 0;
Matches 1768; Conservative 512; Mismatches 777; Indels 184; Gaps 44;
Qy 14 PPTPASFCSHGSDPLNSSYEQLFHLYGLDSSRIEAIKPTCTPFQDMLDNCALDKOSAIGH 73
Db 22 PLLPHDDEKHS- ---NLVEQATRHFGLSRDKIENVLPCTSEFCQDVICDAVDDRRAIGH 77
Qy 74 AVYDPTDIDISRFAALANKEIVNQTPALRAFAFTSDSGTSQVILKDSVFSWMCWSSSS 133
Db 78 VYIDPTVDIQRLLAAMKEVVRQTPILRTGFTSETGDSFQIVLKEG-CLPMMYATCLG 136
Qy 134 SPDEVVRDEAAAAAGSPRCNREVLLEDMTQKCOLVWTFESHALVDVTFQOVLRSVFAAY 193
Db 137 MKGAVIQDEAVAAMTGPCNRYVLEDPSKORLLIWFESHALVDVTFQOVLRSVFAAY 196
```

QY 194 KHEKTHREPPESSDATDTSOSVSVSMSCEDNAVSAATHETHWOTHLNDLNASVPHLSL 253
 Db 197 ---DGRVECPRIK-----DTEHVS-----REFWOHPFEGDLASVPLLP 233
 QY 254 HLMVNPPTTAEHRTTF--PLSQKALSNAICRTALSILLSRYTHSDEALFCAYTEQS-L 310
 Db 234 HLTVCNPNARAHEHLSYTGVP--QRKWSHTSICRAALAVLLSRFTTHSSSEALFGVVTQOSH 292
 QY 311 PFDKHYLADGYOTVAPLRVHCOSNLRASVDWDAISSVDDRLGHLAPGLRDRINTGNG 370
 Db 293 SEDQRRSIDGPARTVPIRLVLCAPDQYVSDVIGAITAHEHAMRGFEHAGLRNIRRTGDDG 352
 QY 371 SAACDFQTVLLVTDGSHNNINGFLQOITESSHPMPCNNRALLLHCOMESSGALLVAY 430
 Db 353 SAACGFQTVLLVTDGAPKTGSLVHSVESDRMPCANRALLDDCOMAGNSASLVA 412
 QY 431 DHNVDSLOTRRLQOFGHLIKLOSPLDSSMAEVLNMTEDYRAEIESMNSQPLEVQDT 490
 Db 413 DHNVDPROMSRLQLGLYLQFHHVLDPLVRELDVYTAEDCAETKWSERLTWQDA 472
 QY 491 LIHHEMLKAVSHSPTKTAIQAWDGDWYSELNDVSSRLAVHIKSLGURAQAQIIPVFEK 550
 Db 473 LIHDTISKWAAGDPNKAFAWMDGWTYAELDNLSRLAVYIQSLDLRPGQAILPLCFEK 532
 QY 551 SKWVIASMLAVIKSGNAFTLIDPNDPPARTAQVVTQTRATVALTSKLHRETVOKLGRCV 610
 Db 533 SKWVATILAVLKVGRAFTLIDPCPSARMAQVCOQTSAIVALTSKLHNTILRSVSRCI 592
 QY 611 VVDELLQSVASDD-FSSLTKSODLAVIPTSSTGDPKGIMTEHRAFSKALKFGASL 669
 Db 593 VVDDLLRSFLPHADGRKATVKPQDLAYVIFTSGSTGEPKGMTEHGRFVSCAMKFGPAL 652
 QY 670 GINSDTRALQFTHAFACILLEIMTTLINGCVCIPSDDDRMNSIPFINRYNWNMAT 729
 Db 653 GMDHTRALQFASVAFACILVEVYALMHGVCVCIPSDDDRLNNVPEFIKRAQVNWVILT 712
 QY 730 PSYNGTSPEDVPGLATLVLGEOMSSVNAIWAPKILQLLNGYQSSSSSTCF---ASN 786
 Db 713 PSYIGTQPEDEVPGQLTVLVEGPISASIRDWTASOVRLLNAYQSSSSTCMSTVTEVSPL 772
 QY 787 STEPNNNGRAVGAHSWIDPNDINRLVPIGAVGELVITESPGIARDYIVPPPPKSPFFTD 846
 Db 773 SLEPNNIGRAVGARSWIIDPEDEPLAPICIGELVIESPGIARDYIIAPPKSPPLLA 832
 QY 847 IPSWYPANTPPDGAKYRTGDLARYASDGSIVCLGRIDSQVKRGORVELGALETHLRQ 906
 Db 833 PPWYPAGKLSNAFKYKTGDLVRYGPDGTIVCLGRKDSQVKRGORVEISAVEASLRQ 892
 QY 907 MPDOLTIVBATKRSOSANSTSLIAFLIGSSYFCN-----RPSDAHILDHDATKATNIK 960
 Db 893 LPSDIMPVAEAIKRSOSSGTVLTAFLIGSSKSGDGNHLSAADAVILDHGATNEINAK 952
 QY 961 LEQVLPKHSIPSYICMLPRATGKIDRRRLRINGKOTLDKQTOGAIVQ-----QAPA 1015
 Db 953 LQOILPOHSPVSYIIHNNLPRATGKADRKRLRSIASKLLGELSONVTQSPTKEKHDA 1012
 QY 1016 PIPVAFADTAALKHLSIWVOSLIDPATVNVGATFELGGNSITAIKVMNARSVMOLKVS 1075
 Db 1013 -----TGIEVKLELWFLSLNLPNSQDVGASFFDLGGNSIIAIKVMNARSAGIALKVS 1067
 QY 1076 NIYQHTPLAGISAVVKGDPDLSYTLIPKSTHEGPEQSYSGRGLWFLDQLDVGLWLYLPY 1135
 Db 1068 DIFQNPFTLAGLVIGRDPAPYNIPTTAYSGPVEQSFAGRLWFLDQELDALWLLPY 1127
 QY 1136 AVMRGPVNVDAALRALAALQORHETLRTTFEDQGVGVQIVVHEKUSEMKVJDLGCS-D 1194
 Db 1128 AVMRGPLHIDATIALIAIQORHETLRTTFEDQGVGVQIVVHASPISDLRIIDVSGDRN 1187
 QY 1195 LDPFEVINOEQTPPFLNLSSEAGWRATLLRLGEDDHILTIYMHIIISDGWSIDVLRDLNQ 1254
 Db 1188 SDYQLQLHQBQTPPFLACQAGRWSLIRLGEDDHILSIYMHIIISDGWSIDVLRDLN 1247
 QY 1255 LYSALKKSDKPLSALTPLPIQYSDPAKWOK--DQFIEQEKOLNYWKOLKDSPPAKIPT 1312

Db 1248 FYSALRGS-DPLSVVSPPLHYRDFSVWQKQVEQETHEHEROLEYVWVQLADSSAAEFLT 1306
 QY 1313 DFARPALLSGDAGCVHVTIDGELYSLAFACNEHNTTSFVVLAAAFRAHVELTAVEDAV 1372
 Db 1307 DFPNRLILSGAGSVVPVIEGELYERLOEFCQVEQMTFPAVLGAFRAHYRLTGAEDSI 1366
 QY 1373 IGTPIANRNRPELEDIIIGCFVNTQCMRINIDHDTFTGLINQVKATTTAAAFENEDIPPER 1432
 Db 1367 IGTPIANRNRQLENNMIGFFVNTQCMRITVDGDDTFESLVRQVRTTATAAEFHODVPER 1426
 QY 1433 VVSALQPSRDLSTPTLAQLIPAVHSQKDLGRFKQGLSVSPVPSKATRTDMEPHLPOE 1492
 Db 1427 VVTALLPRSDLSRNLPAQLTFALHSQODLGFLEGLVABPVSNNKVTREDFVEHLEQ 1486
 QY 1493 TDSLKGSVNFADLFCMETVENVVRPEILNRGLSSRTSPVSLPILPDGIVTLEKLDVL 1552
 Db 1487 AGRLSGNVAFADLFCPETISNVVAIFQILRQGRIRQPTPIAVLPDGLDADRAMGELL 1546
 QY 1553 NVKHVDYPRSSSLADVFQTSAYPDSLVAVDSSCRITYTTELDRQSDILAGMLRRSRMPA 1612
 Db 1547 EIEKAEYPRSSVVDVFRKQVAAPHAFVAVDSASRLTYADLDRSQDQATATLWGRNTA 1606
 QY 1613 ETLVAVAPRSCETIVAFGVLKANLAYLPDVRSPSARVODIILSGLSGPTIVLIGHDTA 1672
 Db 1607 ETLGVILAPRSCQVWAILGILKANLAYLPDVNCPTARLQTLILNHRHLVLLGSNAT 1666
 QY 1673 PPDIEVNVFVIRDALND-----SNADGF----- 1698
 Db 1667 TPDVQIPDVELVRLSIDLRPINOAKLNGHTKSNYSKPNGYTHLKGYSLNLSYKONG 1726
 QY 1699 -----EVTEHD-----STKPSATSLAVLYLTSGSTGRPKGYMIEHRV 1735
 Db 1727 YAQLNHRERNYLDLNGHLLNSDITTSGPSATSLAVYIFTSSTGTPKPKGVMEHRS 1786
 QY 1736 IIRTVTSG-CIPNYPSETRAHMATIAFDGASVYIYALLGRTLVCVDTMTTLDAALK 1794
 Db 1787 IIRLAKNRIISRPSPVAKVAHLSNIAFDAATWEMFAALLGGFLVCIDVMTTLDSKTL 1846
 QY 1795 DVFFREHVNAASHVTSQDVPVLRPRRLSRTLMLFFFLVVTDS--TAPDALDAQGLYQV 1852
 Db 1847 AAFAREQINAAITPALLKQCLANIPITLGR---LSALVIGGDRDLGDQDAIAAHALV-GA 1902
 QY 1853 QCVNGYPTENGVMSTIYPIIDSTESFINGVPIGRALNNSGAYVVDPEQOLVIGVMGELY 1912
 Db 1903 GVTNAYGPTENGVIITNTKNDSTINGVPIGCAISNGAYITDPDQQLVPPGVMGELY 1962
 QY 1913 VTGDLARGYSDKALDENRFRVHTVNDQTVKAYRTGDRVRYIGDGLIEFFGRMDTQFKI 1972
 Db 1963 VTGDLARGYTDPALDAGRFRVQIMNDKAVRATRGDRARYVGDGQIEFFGRMDQVKI 2022
 QY 1973 RGNRIESAETEAALLRDSSVYRDAAVVLQONEDQAPILGFWADHDHSENDKQGSANOVE 2032
 Db 2023 RGHRIEPAEVERAILDQSDARDAVVYIRHQEGEPEWGVFATHGDHSA-EQEEADDOVE 2081
 QY 2033 GWQDHFEESGMYSDIGEIDPSTIGSKGTSMDGSDIDFDEHMEWJGETTRTLHDNRSL 2092
 Db 2082 GWKDFESNTYADMDTIGQSAIGNDTGWTSMYDGSINKAEQWELDDTMTLLDQOAP 2141
 QY 2093 GNVLEITGSGMTILNLDLSRLSVGLPESRSAAAFVKNKATESIPSLAGKAKVQVGTATD 2152
 Db 2142 GHVLEITGSGMVLNLAGLQSYVGLPESRSATFVTKAINSTPALAGAEVHVGTATD 2201
 QY 2153 IGOVDDLHPDLVVLNSVIOYFPPSEYLAETADTLIHLPNVORIFFGGRVSOATNEHFLAA 2212
 Db 2202 INRLRLPDLVVLNSVVOYFPTPEYLVVSVLVRIPGVKRVVFGDIRSHATNRHFLAA 2261
 QY 2213 RALHTLGLKATKDDVRQKMAELEDMEELLEVPAFTS-LKDRPGLVHEVHEILPKNMEA 2271
 Db 2262 RALHSIGKATKADIRQKMTKEMEREELVDPAPFTALLOGLQADRIKHVELILPKNMRA 2321
 QY 2272 VNELSARYAAVVHVHVG-SLGDDELVLVPEKDDWIDFQANOLNOKSLGDL-L-KSSDAIMA 2329

Db 2322 TNELSARYTAVIHVGRPEEGRVYPIQVNDWIDFQASRIDRRALLRLQSRADAATVA 2381
 QY 2330 VSKIFEITAFERQVVASLNSIDE-----WOLSTIRSSAEGDSSLSVPDIFRIA 2379
 Db 2382 VSNIPYSKTIVERHVESLDDNNRTHRAPDGAAW-ISAVESKAERCTSLSVTLVQLG 2440
 QY 2380 GEAGFRVEVSARQSQNGALDAVPHH-----CCSOGRTLVNPPD-DHHLRGSDLTNRPL 2434
 Db 2441 EEAGFRVEVSARQSQNGALDAVPHH-----CCSOGRTLVNPPD-DHHLRGSDLTNRPL 2500
 QY 2435 ORLQNRRTAIEVRRLSLLPSYMLPSNIVLDMKPLNANGVKDRKELSRRAKVVPKQOT 2494
 Db 2501 QRLQSRRTASQIREQKAVLPSYMLPSNIVLDMKPLNANGVKDRKELSRRAKVVPKQOT 2560
 QY 2495 AAPLPTFTPISEV-----EVILCEBATEVFGMKVDITDFFHNLGSHSLATKLISRIDQR 2548
 Db 2561 A-----PAKPVKQVDPVNLAEALICEEFAEVLGMEGVNDHFFGLGSHSLATKLISRIDQR 2617
 QY 2549 LKVRITVKDVPDHPFADLASVIROGLGLOQ---PVSDGQG---QDRSAHMAPRTETAIL 2603
 Db 2618 LKVRITVKDVPDHPFADLASVIROGLGLOQ---PVSDGQG---QDRSAHMAPRTETAIL 2677
 QY 2604 CDEFKVLGFGVIGITDNFDFLGGHSLMATKLAIRIGHRLDTTVSVKDVDFHPVLPQLATA 2663
 Db 2678 CKFADVLGVESVATDSFDFLGGHSLMATKLAIRIGHRLDTTVSVKDVDFHPVLPQLATA 2737
 QY 2664 LDNLVQSKTNEIVGGREMAEYSPFOLLETPPEEFMASEIKPOLELO-EIIDIYPTQM 2722
 Db 2738 I-RLTOAKGHEATNGVQIANDAPPOLISVEDPEIFVQREIAPOLCSPETILDVVPATQM 2796
 QY 2733 QKAFLEDTHTARPREVFYIDPSTSEPDAGLTKACESLVNHLIDIFRTVFAEASGELY 2782
 Db 2797 QRVFLNVPVTKPRSPPTFHIDFP--PDADCASLMRACASLAKHFDIFRTVFAEASGELY 2854
 QY 2783 QVWLSCLDPIQVIEEDNINATNEFDEFAKEPVRLGHPLIRETII-KQTKSMRVIMR 2841
 Db 2855 QVWLSCLDPIQVIEEDNINATNEFDEFAKEPVRLGHPLIRETII-KQTKSMRVIMR 2914
 QY 2842 ISHALYDGLSLHVVYKRLHLYNGRSLPLPPHOFQSRMYQYADGREGHGFWRDVIQNTPM 2901
 Db 2915 LSHALYDGLSLHVVYKRLHLYNGRSLPLPPHOFQSRMYQYADGREGHGFWRDVIQNTPM 2974
 QY 2902 TILSDTIVVDGNATCK-----ALHLSKIVNIPSOVLGRSGNSNITQATVFNAA 2949
 Db 2975 -----TWIKGNNTTPPPPPQOQSTPSGAHHASKVVTIQA--NTDSRITRATIFTA 3026
 QY 2950 CALVLSRESKSDVGVGRIVSGRGLPVEYODIVGCTNAVFPVRA---HIESSDYNOLLH 3006
 Db 3027 CALMLAKEDSDVGVGRIVSGRGLPVEYODIVGCTNAVFPVRA---HIESSDYNOLLH 3086
 QY 3007 DIQDQYLLSLPHETIGFSDLKRNCTDWPE--AITNFSCTIYHNFPEYHPSQEQORVEM 3064
 Db 3087 EMQEQYLLSLPHETIGFSDLKRNCTDWPE--AITNFSCTIYHNFPEYHPSQEQORVEM 3146
 QY 3065 GVLTKEVNIEMDEPLYDLAAGEVDPDAGLKVTVIAKTLQFGRKRVREHLEEVSKTFEG 3124
 Db 3147 GVLTKEVNIEMDEPLYDLAAGEVDPDAGLKVTVIAKTLQFGRKRVREHLEEVSKTFEG 3206
 QY 3125 L 3125
 Db 3207 L 3207

RESULT 3

AAR44929

ID AAR44929 standard; Protein; 15281 AA.

XX AAR44929;

XX AAR44929;

DT 08-JUL-1994 (first entry)

DE T. niveum Cyclosporin synthetase.

XX

Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
 T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase.
 Tolypocladium niveum.
 EP578616-A.
 12-JAN-1994.
 05-JUL-1993; 93EP-0810474.
 09-JUL-1992; 92AT-0001403.
 08-MAR-1993; 93AT-0000437.
 29-APR-1993; 93CH-0001310.
 04-MAY-1993; 93CH-0001375.
 (SANO) SANDOZ LTD.
 (SANO) SANDOZ PATENT GMBH.
 (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
 Leitner E, Schneider E, Schoergendorfer K, Weber G;
 WPI: 1994-010432/02.
 N-PSDB; AAQ54386.
 Isolated DNA sequence - which codes for enzyme having cyclosporin
 synthetase like activity
 Claim 1; Page 41-84; 93pp; English.
 This sequence represents an enzyme which has cyclosporin synthetase-
 like activity. This sequence was isolated from Tolypocladium niveum
 (formerly known as T. inflatum GAMS). This enzyme catalyses the
 peptide biosynthesis of cyclosporins and structurally related
 molecules. This sequence may be used for the production of
 cyclosporin by transforming a vector containing this sequence in
 to a recombinant host. This allows effective production of anti-
 biotic cyclosporin or its derivatives.
 Sequence 15281 AA;
 Query Match 31.0%; Score 5000; DB 15; Length 115281;
 Best Local Similarity 32.4%; Pred. No. 0;
 Matches 1233; Conservative 440; Mismatches 867; Indels 1264; Gaps 66;
 QY 482 SOPLEVQDTLIHHEMLKAVSHSPTKTAQWMDWTYSBELDNVSRSLAVHKSGLRAQQ 541
 Db 11581 NQPDYPRASVIDVFEQVASTPKSTAVIDASSQLTYTDELSSQLATWLR-QVTPPE 11639
 QY 542 AIIPIYFEKSKVVIASMLAVLKSNGNAFTLIDPNDPPARTAQVVTQT--RATVALTS--- 595
 Db 11640 ELVGLAPRSCETIIAFLGIIKANLAYLPDYNAPAGRIETILSSLPGNRLILGSDTQA 11699
 QY 596 -KLHRETQKLVGRVVDDELLOSVS-ASDFFSLTSQDLAYVIFSGSGDPAKIMI 553
 Db 11700 VKLHANSV-----REFRISDALVESGSPPTSTRAQSLAYVMTSGSGVPGKGVW 11754
 QY 654 EHRAFSSCALKFGASLGINSDFRALQ-----FGTHAFAGCLLEIMTLLNGG---CV 702
 Db 11755 EHRGIT-----RLVKNNSNVVAKQAAAAHAHLSNTAFDASSWEIYAPLNGGTVVCI 11806
 QY 703 CIPSDDDRMNSIPSFIRYNVNMWMPATPSYMG---TFSPEDVPLGLATLVGEQMSVVN 759
 Db 11807 DYYTTID-IKALEAVFKQHHRGAMLPALLKQCLVSAPTMISSLEILLPAAGDRSSQ-D 11864
 QY 760 AIWAPK---QLLNGYQSESSSICFASNM-STEPNNMGAVG-----AHSWIDPDNDIN 810
 Db 11865 AILARRVGSVYINAYGPTNTVLTSTIHGIGNEAFSGVPTGNVAVNSGAFVMDON--Q 11922
 QY 811 RLVPITGAVGELVIESPGIARDYIVPPPEKSPFFDIPSWYPANTF---PDG--AKLYR 864
 Db 11923 QLVSGAVIGELVVTGDLARGY-----TD--SKLRVDRFIYITLIDGNRVRAYR 11968

| | | | |
|----|-------|--|-------|
| Db | 14071 | LDGVDEPVLSTVSTRFDLEFHAFOEADRLNGSVMFATDLFQPETIQGFVAVVEVLQRG | 14130 |
| QY | 2577 | LOQPV----- | 2582 |
| Db | 14131 | LEQOSPIATMPLAEGIAQLRDAGALOMPKSDYPNALSIVDFQOQAMASPSTVAVTDS | 14190 |
| QY | 2583 | ----- | 2582 |
| Db | 14191 | SKLTYAELDRSLDQAASVLRROOLPAETMVAVLAPRSCETIIAFLAILKANLAYMPLDVN | 14250 |
| QY | 2583 | ----- | 2582 |
| Db | 14251 | TPSARMEALISSVPGRRLLILVSGSVRHADINVPNAKTMLSIVTGTDAIGTPEPLVVRP | 14310 |
| QY | 2583 | ----- | 2582 |
| Db | 14311 | SATSLAYVIFTSGTGKPGVMVEHRAIMRLVKDSNVVTHMPPTMAHVTHIAFDVSLF | 14370 |
| QY | 2583 | ----- | 2582 |
| Db | 14371 | EMCATLLNGTLVLCIDYLTLLDSTMLETREOVRAAIFPPALLRQCLVNMPPDAIGMLE | 14430 |
| QY | 2583 | ----- | 2582 |
| Db | 14431 | AYVAGDRHRSRDARATOMLAGPRVYNAYPTENAILSTIYNIDKHDPVYVPIGSAVS | 14490 |
| QY | 2583 | ----- | 2587 |
| Db | 14491 | NSGAYVMDRNQOLLPPGYMGLVVTGEGVARGYTDASLDTDRFVTVTDGQRAYRTGD | 14550 |
| QY | 2588 | ----- | 2587 |
| Db | 14551 | RVRYRPGQIEFFGRLDQQAQIRGHRVELGEVHALLSENSTVDAVVLRMTWEDPOL | 14610 |
| QY | 2588 | ----- | 2587 |
| Db | 14611 | VAFTVTDHEYRSGSSNEEDPYATQAAGDMRRLSLDPPYVWVSRVTILRMPNLNANK | 14670 |
| QY | 2588 | ----- | 2629 |
| Db | 14671 | VDKDLARRAQMTPTASSGGPVHVAPRNETEAICDEFETILGVKGITDNFELGGHSL | 14730 |
| QY | 2630 | MATKLAVRIGHRLDITVSKVDHPVLPOLALDNLV-----OSKTNEIVGGREMAE | 2683 |
| Db | 14731 | LATKLAARSRMRGRISVKOLFDDPVVPSLAKLEQQOQGFSGEDESSTVGIV----- | 14783 |
| QY | 2684 | YSPFOLLFTDEPEEFMASIKPOLEL-QEIIODIYVSTOMAKAFLDHTTARPRFPVPY | 2742 |
| Db | 14784 | --PFOLLPAEMSRELIQDVVQIENGHSTPLDMTPATQTOFFLHDKATGHPATPPLFS | 14841 |
| QY | 2743 | IDPSTSEPDAAGLIKACESLVNHLIDIFRTVFAASGELYQVVLSCLDLPQIVETDNI | 2802 |
| Db | 14842 | LDFPETA--DCRRLASACAALVQHFDIFRTVFSRGGRYQVLAHLDPVEVIEQEQL | 14899 |
| QY | 2803 | NATNEFDDEFAKEPVRGLHPLIRFTIITKQT-KSMRVIMRISHALYDGLSLHHVVKLHM | 2861 |
| Db | 14900 | DEVALALHADQQLPURLGRAMRLIARLKPGRKMRVLMSHSLYDGLSLHHVNALHA | 14959 |
| QY | 2862 | LYNGRSLPLPHPSRYMOTADRGSGHGFVRDVTQNTPTILSDDT-VVDGNDATCKAL | 2920 |
| Db | 14960 | LYSDKHLAQAPKFLGYMHMAWRRAEAGYNFWSILQGSMTSLKRSVGALEAMTPSAGTW | 15019 |
| QY | 2921 | HLKSVINTPSQVLRSSNITITQATVFNACALVLSRESKDVFGFRIVSGROGLPVEVQ | 2980 |
| Db | 15020 | QTSKSIPTPPAALK--NGITQATLFTAAVLSLLAKHTKSTDVDFGRVVSGRQDLSINQ | 15076 |
| QY | 2981 | DIVGCTNAPVRAHI-ESSDYNQLLHDIDQOYLLSLPHETIGFSDLRKNCNDWEAINT | 3039 |
| Db | 15077 | DIVGPCINVPVRVIDEGDGMGLLRIADQOYTSFRHETLGLQEVKENCNDWDATKE | 15136 |
| QY | 3040 | FSCCITYHNFVEHPESQFEQKRVEM----- | 3064 |
| Db | 15137 | FSCCITAFQNLNLHPEAIEGQIRLEGLPAKDQARQANGHAPNGTNGTNGTNGANCT | 15196 |

QY

3065

-----GVLTKFVNIEM-----

DEPLYDLAIAAGEVEPDGAGLKVTVIK

3102

Db

15197

NGTNGTNGHANGNSGNGRDSNVVSAAGDQAPVHDDIVGIPEDPGS-VKIGIGAS

15255

QY

3103

TOLFGRKRVRHLLLEVKSTFEGLN

3126

Db

15256

ROIIGKVVGSMNLCELTMLALS

15279

RESULT 4

AAB07581

ID

AAB07581

standard; Protein; 2841 AA.

XX

AAB07581;

AC

AAB07581;

XX

20-OCT-2000

(first entry)

XX

Protein encoded by the bleomycin (BLM) gene cluster ORF31.

XX

BLM gene cluster; bleomycin gene cluster; polyketide metabolite;

KW

bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;

KW

thiazoline; bithiazoline; microbial metabolite; sugar.

XX

Streptomyces verticillus.

XX

WO2000040704-A1.

XX

13-JUL-2000.

XX

06-JAN-2000; 2000WO-US00445.

XX

06-JAN-1999; 99US-0115435.

PR

05-FEB-1999; 99US-0118848.

PR

05-JAN-2000; 2000US-0477962.

XX

(REGC) UNIV CALIFORNIA.

XX

Shen B, bu L, Sanchez C, Chen M, Edwards DJ;

XX

WPI: 2000-465974/40.

DR

N-PSDB; AAA58472.

XX

New bleomycin gene cluster components useful for peptide and/or

PT

polyketide metabolites, especially bleomycin, production and for

PT

chemically modifying biological molecules -

XX

Disclosure; Page 138-153; 162pp; English.

XX

AAB07580-89 represent proteins encoded by open reading frames (ORFs)

CC

8 to 30 of the BLM (bleomycin) gene cluster. The proteins encoded

CC

by the gene cluster are useful for producing peptides and/or polyketide

CC

metabolites, especially bleomycin or bleomycin analogues. They are

CC

also useful for chemically modifying biological molecules to produce

CC

branched methyl groups, and for coupling amino acids and fatty

CC

acids. They may be reacted with an apo-carrier protein and coenzyme A

CC

to produce a holo-carrier protein. The BLM gene cluster or catalytic

CC

domains can be used individually or collectively to produce

CC

thiazolidine, thiazoline, bithiazoline and bithiazoline-containing

CC

microbial metabolites. The BLM gene cluster may also be used to produce

CC

sugars.

XX

SQ

Sequence

2841 AA;

Query Match

13.3%;

Score

2146.5;

DB

21;

Length

2841;

Best Local Similarity

28.9%;

Pred. No.

3.9e-163;

Matches

722;

Conservative

394;

Mismatches

1025;

Indels

357;

Gaps

91;

QY

235

FWOHTLNDLNASVFFHLSDLHMPNPNTTAAE--HRTTFPLSOKALS-----NSAIC

283

Db

539

YMKRAL-DCAPSVLRPLRMDH---PRPAVQSERGETGVGFALPDALVAALKLREGQCATLF

594

QY

284

RT---ALSTLLSRYSDEALFGAVTEQSLPDKHYLADGTYQTVPAPLRVHCQSNLRASD

340

Db 595 MTLGAFQVLLARHAGQEDIVVGPAAAGRTTETEPV-GFVNTLPLRAICAGLSFRD 653
 QY 341 VMDALSSYDDRLGHILA---PF-----GLDIRNTGNGSAACDFQTVLLVTDGSH----- 387
 Db 654 LLDQVR--EALGAFAHQDLPEALVEALAPERDLGHNPVQVTFQ---LLGTPAARPLDI 709
 QY 388 -----VNGINGCF-----LOQITESHFWPCNNRALLHC-----OMESSGALLV 427
 Db 710 GTEVERYPQBAVSQFSLDIKRAODGSY-----RGILNYCPDLFDRRME-----VLV 759
 QY 428 AYV-----DHNVIDSLQTLTLLQOFGHLIKCLOSPDLSSMAEVLNMT 471
 Db 760 GHYLLTLLGAAADPCRPGLPLSDGAERLRLDGFGRDAAYAGP----- 805
 QY 472 YDRAIESWNSOPILEVQDTLHHEMLKAVSHSPKTAIAQWDGWTYSELNDSVSRSAVH 531
 Db 806 -----GSPV-----ERFAEVARTAPARAVTCGATLTTLFAELNDRVERLAQA 847
 QY 532 IKSIGLRAQQAIIPIVPEKSKWVIASMLAVLKSGNATFLIDPNDPPARTAOVVTQRTATV 591
 Db 848 LLGAGV-TRETPVAVRPRSDSVVALLAVNRAGGVVPLDPDPADRTAYILDDTAASV 906
 QY 592 ALTKLHRETQKLVGRVW-----VDDELQSVASDDFSLTKSQDLAVVIFTSSTG 646
 Db 907 VITRDL-----PALPGRLHVDPRAADGLVPAPRIDPD-----QAAYVITSGSTG 953
 QY 647 DPKIMTEHRAFSCALKFKGASL-----GI-----NSDTRALQFGTHAFACILLEI 692
 Db 954 APKGVVVRHRSNLHLSALQATFLGHPDYPLAGADGVPPGDAKLRTLTAPTFTDASMEQL 1013
 QY 693 MTTLLINGCVCIPSDDDR-MNSIPSFNRYNVMNMATPSVM-----GTFSPEDVPGLA 745
 Db 1014 SWMLAGHELFTVPEVDRDPSALVRFRHREHVIDTTSQLELLVSHGLDGEWAP--- 1070
 QY 746 TLVLVGQMSSSVNAIWAPE-----KLQLLNGYGOSESSICFASNMSTENPMNGRAYGA- 799
 Db 1071 SMWVVGGEAVSP--SLARTLRDQRTRCFNLYGTEATVATCDHLS-DEADV-PVIGTP 1126
 QY 800 --HSWIDPNDINRPIGAVGELVIESPGIARDYIYVPPPEKSPFTDIPSWYPANTFP 857
 Db 1127 LPHTHVRLDRLRPVPGVAGETIYLGCTGLARGYLRNPALTARRFVAD-----PYP 1178
 QY 858 D--CAKLYRTGLDARYADSGISVCLGRIDSOVKIRGORVELGATETHL-RQMPDOLTI 914
 Db 1179 DTGSRLYRTGDRARWPDGTLEVLGRTDDQIKIRGFRVPEGEIEAVLTHHPAVKEAAV 1238
 QY 915 VEATKRSQSANSTSLIAFLIGSSYFGNRPDSADHILDHDKATKAIKLEQVLPKHSIPSFY 974
 Db 1239 DDAHAR-----LVAYVTLAEGGAGPTD-----VRRFAQGRLPAMHWPASV 1279
 QY 975 ICMLELPRATGKIDRRRL--RINGKIDLKQGTGAIVQAPAPIPVFADTAAKLHSIW 1032
 Db 1280 VLEALPLTNGKLDRLARLPAPAGRPELD-----VRFVAPRDMVEEWAQ---VMC 1328
 QY 1033 QSLGIDPATVNGATFPELGNSITATK-MVNMAVSGBMDLKNYNIQHETLAGISAVVK 1091
 Db 1329 AVLQVD--RVGVHDDDFELGSHLLVVQVMTIRKLGLVREPLREUFAAATVEELAAVR 1386
 QY 1092 -----GDPLSYTLIPKSTHEGPVEQSYSGGRIMFLDQDLVGLSWLYPIYAVMRGPVN 1144
 Db 1387 AARTEGLGRCAAPPLGPVD-RSGPLPLSFAQRLWYLDQLAPDSVSNMCDAYVRGPDL 1445
 QY 1145 VDALLRALAALQORHETLRTTFEDQGVGVQIVH-----EKUSEMKVYIDLGC-SDLD 1197
 Db 1446 LDALLRALRTLVERHETLRTAFYERDGVPHQVVSAPDAPARAAAEVVRTEAAGRTDEAV 1505
 QY 1198 FEVLNBQOTTPEPNLSSAGWRATLLRLGEDHLLITVMHHIISDGHSIDVLRDLNQLYS 1257
 Db 1506 RDLVAEARTPPFADGALMRVAVARLADDDHVLVVTTHHIVSDGWSVDLDELGRLYR 1565
 QY 1258 AALKDSKOPLSALTPLIQYSDFAKWKDOFI--EQEKQLNYWKQKDSSPA--KIPTD 1313

Db 1566 EHV--TGDP-AGLPDLDVOYADFAVWQSRWMTGVPVREEHLYWKRAL-DGAPSVLRLPAD 1621
 QY 1314 FARPALLSGDAGCVHVTIDGELYQSLRAFNCNEHNTTSFWLLAAFAAAHYRLTAYEDAVI 1373
 Db 1622 HPRAVSORGETVEFPPLPVARLEALCALCREQVTLFMAFGAFVOLLARYSGQDDVVV 1681
 QY 1374 GTPIANNRPELEDIIGCFVNTQCMRINIDHHDFTGTLINOVKATTTAAFENEDIPFERV 1433
 Db 1682 GVPANTRABTEPLVGFFVNTLPRVACSPELSFRALLDRVREAAALGAFHQDLPPEAL 1741
 QY 1434 VSALQPGSRDSSPLAQLIFAVHSQDKLGRFKQGLSVESVPVPSKAT-REDMEHFLFOE 1492
 Db 1742 VEALAP-ERDLGHHPLVQVTFQLLDAPD-ERLVLHGTDCVSLGFGGVTSRDLSDLVVSG 1799
 QY 1493 TDSLKGSVNADELFKMETVENNVVRFEEILNGLQSSRTPVSTLPLTDGVTLEKLDVL 1552
 Db 1800 RRGKRCVLYTCPLDFDRPRMEVLAGHYLLTLLGAADDPGLRGDPLPSDDVERLRLGGS 1859
 QY 1553 NVKHVDYPRESSLADVTQTSVAYPDSLAVVSSCRITYTBTDRQSDILAGLRRRRSMPA 1612
 Db 1860 RPYLPAPGAETVPDAFAAQVRATPDAPALVHGDSTLTFAELDTRTALAVRLRRCGVAA 1919
 QY 1613 ETLVAVFAPRSCETIVAFFGVKLANLAYPLDVRSPSARVQDILSLSGPTIVILGHDTA 1672
 Db 1920 ETPVAVCLPRSADAVALLAVLRAGGVTVVPDPEMPSGRVAHVLDETAAPVVITRDLUPAD 1979
 QY 1673 PPDIEVTNVEFVRIRDALNDSNADGFEVIEHDSHKPSATSLAYVLYTSGTGRPKGVME 1732
 Db 1980 PGRVHLDP-----RQAPAD-DRDPLRLHRDQA-----AYIIFTSGTGAPKGVVR 2025
 QY 1733 HRVI-----IRTVTSCIPNYPSETRMAMHATTAFDGAHYEISALLFGTLVCVDYMT 1786
 Db 2026 HGSYHLHGVRRAEAG-----GPRNVVAHTMTATFD-PSLEQFLWLAVAGTLHVAPEV 2079
 QY 1787 TLDAARALKOVFFREHVN---AASHVTSSQDVP---RVPRLSRLTLMFFLVYTDSTA 1839
 Db 2080 RRDPEALVALVRAAIDVNLVTPSHLTLLEAGLEGRDVPGT-----VLVGGEAV 2130
 QY 1840 PDAL-----DAQGLYOGVQCYNGYPTENGVMSTIYPIDSTESINGVP-IGRALNNSGA 1893
 Db 2131 PAALWRTLRERTG---ATREFNLYGPTTEATVATCDHLSDPAD---VPVIGTLPHTHV 2183
 QY 1894 YVVDPEOQLVIGVWGLVVTGDLARGYSDK-ALDENRRFVHTVNDOT-VKAYRTGDRV 1951
 Db 2184 RVLDRLRPVPGVAGEIYLGCTGLARGYLRNPALTQAQRFVADPYPTPGSRLYRTGDRA 2243
 QY 1952 RYRIGDGLIEFFGRMDTQFKIRNRIESAETEAALLRDSVVRDAAVVQOQNEOQAPETLG 2011
 Db 2244 RWR-PDGTLEYLGRTDQIKIRGERVEPEGEIEAVLTHHPAVKEAAVTV-ATDDGAARLVA 2301
 QY 2012 FVV--ADHDHSENDKQOSANQVEGWQDHFESGMYSIDIGEIDPSTIGSDFKGTSMYDGSQ 2069
 Db 2302 LVVPAPRAPHGDSADGAPDAQVEENAVFEA-THTDAAD---GELTFNKGWNSDLTGAP 2357
 QY 2070 IDDEHHEWLGCTRTLHDNRSLGNVLEIGTGSOMILFNLDLSRLESVGLSPRSAAAFV 2129
 Db 2358 IPAEHMRWYDVTVARLLE-RPAERVLGSGTGLLWMLRPLHVTEYTGDFSRPAYDWL 2416
 QY 2130 NKATESPSLAGKAKVOVGATDITGOVDDLHPLDVLNLSVIOYFPSSEYL-AETADTLIH 2188
 Db 2417 RDGLRRRP--AHRVRLHREATDFTGVRAASTDLVVVNSVQVFPDRAYLDTVLARALDA 2474
 QY 2189 LPNVQRIFFGDVRSQATNEHFLAARAHTLGKNATKDDVRQKMAELDEMEELLEVPAFF 2248
 Db 2475 TADRGVRFVGDVRNLALAPOFYARQALAHAGPAAARAVARAAGEFAAMDGELLVSPAYF 2534
 QY 2249 TSLKDRFPGLVEHVEITLPKMWEAVNELSARYAAVHVVRGSLGDELVLP-VEKDDWIDFQ 2307
 Db 2535 AALAAKSP-RVTGVEILPRGRHNRNMSLYRYDVLHVQ---GDRPAAPAEAYLTWGD-Q 2589
 QY 2308 ANOLNOKSLDILLKSDAAIMAVSKIPFEITAFEROVVASLNSNIDENWQLSTRSSAEGD 2367
 Db 2590 VHDLASISARLGGGPDALLV-----RGVANDRLTRDNELLDAPARTTA--- 2633

Db 2203 A-----ELIRAAALRIEAKKHLHLLDMHHIADGVSRGIFVKELALLYKGEQPEPTLHYK 2258
QY 2874 FSRMYOYTAGRE---SGHGFWRDVQ-----NTPTILSDDTV-VDCNDAT 2916
Db 2259 DRAVQWNEAEQERMEKHEAYWMSVLSGELPELDLPDVARPPVQSFKGDTIRFTGSET 2318
QY 2917 CKALHLSKIVNIPSOVLGRSSNIITQATVFNACALVLSRESKDVVFGRIVSGROGLP 2976
Db 2319 AKA--VEKLL-----AETGTLHMLHVAHFVFLSKISQORDIVIGSVTAGRTNAD 2367
QY 2977 VEYQDIVGCTNAPVRAHI-ESSDYNQLLHDIOQOYLSSLPHETIGFSDLKRNCTDWE 3035
Db 2368 V--QDMPGMEFVNTLALRMEAKEQOTFAELLEAKQTNLSALEHQEYFFEDLV-NQLDLPR 2424
QY 3036 AITN---FSCCITYHNFEVHPESQFQOQVEMGVLTGFKNVEMDEPLYDLAIAAGEVDPG 3092
Db 2425 DMSRNLFFNMVVTEN---PDK--BELTLQNLISIPY-EAHQCTSKFDLTIGGTFDENG 2477
QY 3093 AGLKVTVIAKTQFLGRKRVEHLLLEEVSKTFEGL 3125
Db 2478 IGLQLEY--ATDLFAKETAEKWEYVRLLLKAV 2508

RESULT 6

AAU36277
ID AAU36277 standard; Protein; 2448 AA.

XX AC AAU36277;

XX DT 14-FEB-2002 (first entry)

XX Pseudomonas aeruginosa cellular proliferation protein #267.

XX Antisense: prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

XX OS Pseudomonas aeruginosa.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-0509180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

XX N-PSDB; AAS541136.

XX New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 11870; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

XX prokaryotic cellular proliferation, their use in identifying the

XX genes, their use in the discovery of novel antibiotics, the essential

XX genes, themselves and the encoded proteins. The prokaryotes used are

XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

XX invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 2448 AA;

Query Match 10.6%; Score 1713.5; DB 22; Length 2448;

Best Local Similarity 28.6%; Pred. No. 3.8e-128;

Matches 586; Conservative 327; Mismatches 750; Indels 385; Gaps 76;

QY 430 YDHNVIDSLQTTLLQOFGHLIK-CLQSPDLDSMAEVLNMTDYRAE-IESWNSQPLE- 486

Db 441 YATDLEDASTVERLAGHWRNLLRGIVANP--RQRLGELPLLDAPERRQTLSENNPAQREC 498

QY 487 -VODTLIHHEMLKAVSHSPKTAIQAWDGDWMTYSELNWSRLAVHKLSGLRAQAIIIP 545

Db 499 AVQGTLL-QORFEQARQPOQAVAILDEQRLSYGELNANRLAHCLIAIRGVGAD--VP 554

QY 546 V--YEKSKWVIASIAVLKSGNAFTLIDPNDPPARTAOQVVTOTRATVALTSKLHRTVQ 603

Db 555 VGLALERSLDMLVGLLAILAKAGGAYLPDPAAPERLAHLIDDSGVRLLLTQGHLLERLP 614

QY 604 KLVG-RCVVVDDELQSVASDDFSSLTQSQDLAYVIFTSGTGDKPKGIMIEH-----RAF 658

Db 615 RQAGVEVLADIGLVDCYAESDPLTLLS-ADNLAYVIYTSSTGKPKGTLTHRNALRLF 673

QY 659 SSCALKFGASLGINSOTRALQFQTHAFGACLLIEIMTLINGGCVCV-----PSDDDR 710

Db 674 SATEAWFGED---ERDVTLL-FHSYAFDFSWEIFGALLYGGLVIVPQVWSRSPEDFYR 729

QY 711 MNSIPSFIRYNNVNMNMTPS-----YMGTFSPEDV---PGLATLVLVGGEOMSSSVNAIW 762

Db 730 L-----LCREGVTVLNQTFSAFKQLMAVACSDMATQOPALRYVIFGGGALDLSLRPW 783

QY 763 APKL-----QLNGYQGESSESSICFASNMSTP-----NNMGRVAGAHWSVIDP 806

Db 784 FQREGDRQPOLVNMVYGITETT-----VHVTVRPVSEADLEGGLVSPIGGTIPDLWSVILD 838

QY 807 NDINRLVPGAVGELVIESPGIARDYIVPPPPPEKSPFFTDIPSWYPANTFPDGA--KLYR 864

Db 839 RDLNP-VPRGAVGELYIGRAGLARGYLRRFLSATRF---VP-----NPPFGGAGERLYR 889

QY 865 TGLARYASDGSIVCLGRIDSOVKIRGORVELGAETHLRQOMPDDLTIVVEATKRS-QS 923

Db 890 TGLARFQADGNIEYIGRIDHQVVRGFRLEIEIEAALA-----GLAGVRDAVLAHDG 944

QY 924 ANSTSLTAFLLIGSYFGNRPESDAHILDHDATKAINIKLEQVLRPHSPSYICMLLEPRT 983

Db 945 VGGTQLVGYVVADS-----AEDAERL-----RESLRESLKRHLDPYMWPAHMLLERMPLT 995

QY 984 ATGKIDRRRLRMKGDKILDKOTOGAIVQQA-PAPIPVFADTAALKHSIWVQSIGIDPATV 1042

Db 996 VNGKLDKRAL-----PPDASLSQQAYRAP---GSELEQRIAAINSEILGVE--RV 1041

QY 1043 NVGATFFELGCGNSITAIKVMNMAR-SVGMDBLKVSNIYOHPTLAGISAVVAGDPLSYTLIP 1101

Db 1042 GLDDNFFELGCHSLLATRVISRVREQEQLDASKALFERPVEAFACGLERTDVAVSTIP 1101

QY 1102 KSTHEGFEVQSYQGRFLWFLDQDLVGLWYLIPYAVMRGFPVNVDAURLAALAEQRHET 1161

Db 1102 LADROQPLALSFAQERQWFLWQLEPESAAVHIPSALRLGRDLVDALQSFDSLSVARHET 1161

QY 1162 LRTTF-----EODDGVGVQIVHEKLSSEMKVIDLCGSDLDPPFVLNQEQTTFPNLS 1212

Db 1162 LRTTF-----EODDGVGVQIVHEKLSSEMKVIDLCGSDLDPPFVLNQEQTTFPNLS 1212

Db 1162 LRTFRLEGGRSYQVPAVSVSIEREQFGEGLI-----ERIQAIIVQPFDLE 1210
QY 1213 SEAGWRATLLRGEDDHILITVMHHIISDGSINDVLRDLNQLYSAAKDSKPLSALTP 1272
Db 1211 RGPLRLVNLQLAEDDHVLVLQVHHIVSDGSMQVMVEELVQLY--AAYSQGLDVV--LPA 1267
QY 1273 LPIQYSDPAKWKQD--QFIEDEKOLNWKKOLKSSPA-KIPTDFARPALLSGDAGCVHV 1329
Db 1268 LPIQYADYALMQRSMWAGERKQLAYWTGLLGEQVLELFPDRPRAROSHRAQLGF 1327
QY 1330 TIDGELYSRAFACNEHTTTSFVYLLAAAFRAAHYRLTAVEDAVICTPIANRNRELEDI 1389
Db 1328 ELSRELVEAVRAQAQEGASSEMFLASFOALLRYSCQADIRVGPPIANERNVETERLI 1387
QY 1390 GCFVNTQCMRINIDHDTFTGLINQVATTTAAEFENEDIPERVVSALQPGSRDLSSTPL 1449
Db 1388 GFFVNTQVLKADLDGRMGFDLELAQAQRALAEQAHQDLPEQLEVALQP--ERNASHNPL 1446
QY 1450 AOLIF-----AVHSOKDLGRFQFGLSVVPFSKAY----TREDMEHFLFOETDSIKG 1498
Db 1447 FOVLFNHQSEIRSVTPVQLEDLLEGL-----AWDQTAQFDLTLDIQEDENGIIWA 1498
QY 1499 SVNFADFLKMETVENVVRVFFELIRNGLOSSRTPVSLPLDGTGIVTELEKLDVLNVKHDV 1558
Db 1499 SFDYATDLDEASTVERLAGHWRNLRGIVANPRQRLGELPLDAPERRQTLSEWNPQRE 1558
QY 1559 YPRESSLADYFOTQVAYSPLAVSDSCRLTYTELDRQSDILAGWLRRRSPAPETLIVAV 1618
Db 1559 CAVGCTLOQRFEQARQPOQAVAILIDELRSLYSGELNARANRLAHCLIAICRGVADVPVL 1618
QY 1619 FAPRSCETIVAFVGLKANLAYLPLDVRSRSPARVODILSGLSGPTIVLI--GH--DTAPPD 1675
Db 1619 ALERSLDMVLGLLAILKAGAYLPLDPAAPERLAHLDD--SGVRLLLTOGHLLERLP-- 1675
QY 1676 IEVTNVEFVRIDALNSADGFEVIEHDKTKP-----SATSLAYLYVTSGTRPKGVMI 1731
Db 1676 -ROAGVEVLA1-----DGLVLDGYAESDPLPTLSADNLAYVITSGTRPKGTLL 1725
QY 1732 EHRVIR--VTSICIPNYPSEFMAMHATADGASYEYISALLFGLTLCVVDYMTILD 1789
Db 1726 THRNALRLFSATENWF--GFDERDVTWTLFHSATDFSVWEIFGALLYGRVIVPQWVSRS 1784
QY 1790 ARALKDYFFREHVNAASHVTS-----SQDPLRVPRRLSRTLMFFFLVVVTDSTAP 1840
Db 1785 PEDFYRLCREGVTVLNTPSAFKQLMAVACSADMATQQP--ALRYVIF-----GG 1833
QY 1841 DALDAQGLYQVQCY-----NGYPTENGVMSTIYPIDSTESFNG---VPIGRALN 1889
Db 1834 EALDQSLRPFQRFQDQFQPLQNMVYGITETTVHVTYRPV--SEADLKGGGLVSPIGGTIP 1891
QY 1890 NSGAYVVDPEQQLVGIGVMGELVVTGDLARGYSK--ALDENRFV--HITVNDQTVKAYRT 1947
Db 1892 DLSWYILDRNLNVPVRAVAGELIYGRAGLARGYLRRPCLSATRFVPPNPFPGGAGERLYRT 1951
QY 1948 GDRVYRIGDGLIEFFGRMDTQFKIRGNRIESAEIEAALRDSVSDAAVVAALQONEDQAP 2007
Db 1952 GDLAREQ--ADGNETEYIGRIDHQVKVRGFIELGEIEAALAGLAGVRD--AVVLAHDGVGGT 2009
QY 2008 EILGFVVADHDHSENDKQSANQVEGQDHFEGSMYSDI-----G 2047
Db 2010 QLVGYVVAD-----SAEDAERLRESLRSKRHLRPLDYMVPVPAHMLLERMLPTVNG 2059
QY 2048 EIDPST1-----GSDFKG-----WTSMYDGSQIDFDEMHEWLGTTTTLH 2087
Db 2060 KLRQALPQDASLSQAYRAPSLEQRATAIWAELVGERVGLDD----- 2106
QY 2088 DNKSLGNVLBIGHTGSGMIFNLDSRLSYVGLFSPRSAAAFVNKATESIPSLAGKARQV 2147
Db 2107 -----NFFELG--GHSLLLLMLKERIGD-----TCQATLSISQMLTHASVAE 2146
QY 2148 GTATDQVDDHPLDPLVNLN-----VIOYFSSSEYLAETDLIHLPNWQRIFFGDVR 2201
Db 2147 QAACIEGQARE--SLVPLNGRREGSPFLMFHPS-----FGSVH 2183

QY 2202 SQAT-----NEH-----FLAARAIHTLGNATK--DDVRQKMALEDMEELLVPAFFTSL 2251
Db 2184 CYKTLAMALRDRHPVKGVVCRALLGAGREVPEWDDMVAEYAE----- 2225
QY 2252 KDRFPGLVHEVEIILPKMNEAVNELSAYRYAAVHVHRSGLGDELVLVPEKD----- 2301
Db 2226 -----QLLOHPEGVFNLAGW-----SLGNLAMDVAARLEQRGRQVAF 2264
QY 2302 -DWIDFOA 2308
Db 2265 VGWIDAPA 2272

RESULT 7

AAU33611
ID AAU33611 standard; Protein; 4342 AA.

XX AAU33611;

XX 14-FEB-2002 (first entry)

XX Pseudomonas aeruginosa cellular proliferation protein #55.

XX Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

XX Pseudomonas aeruginosa.

OS WO200170955-A2.

PN 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207272P.

PR 23-OCT-2000; 2000US-242578P.

PR 22-NOV-2000; 2000US-253625P.

PR 27-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS51470.

XX New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 5107; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

Sequence 4342 AA;

| Query Match | 10.4% | Score 1677; | DB 22; | Length 4342; |
|-----------------------|-------|--|------------------|-----------------------|
| Best Local Similarity | 23.6% | pred. No. le-124; | | |
| Matches | 814; | Conservative 498; | Mismatches 1281; | Indels 852; Gaps 135; |
| QY | 65 | LDKQSAIGHAVYDVP-----TDIDISRFALAMKEIVNOTPALRAFAFTSDSGKTSQVIL | 118 | |
| Db | 690 | IDPQS-----AAYNIPGGLRURGELDEALRASQRLVERHEALRFLERDGAALQRIDE | 745 | |
| QY | 119 | KDSFVFSWMCWSSSSSPDEVVRDEAAAA-----SGPCRNFLVLEDMQTK | 164 | |
| Db | 746 | RGEFAWQFVDLAALAE-----HERAAAAQAARAEAAQPFDEKGPLLR--VSLVRLDEQ | 798 | |
| QY | 165 | KCOLVWTFSHALVDVTFQORVL-----SRVFA----- | 191 | |
| Db | 799 | EQHLWTLHHIVADGWSLNLLDEFSLYAEACGGOPADLAPLELHYAEFAAQWQLDA | 858 | |
| QY | 192 | -----AYKHEK--DTHRPETPESSDATDTSQSVSVSMSCEDNAVSAETHWOTHLD | 247 | |
| Db | 859 | EGEARQLAYWRRLGOT--APVLELATDHPRTAQSPAARYSLRVDEALARAIREAL--D | 916 | |
| QY | 243 | LNASVFPHL--SDHLMVNPNTTTAEHRITPP--LSQKALSNSAICRTALSILLRYT---- | 295 | |
| Db | 917 | HEASVEMWLLAAFOALLHRRHSGQETRIGVPSANRQRLQTOGLVGFFINTLVLRGTPRAR | 976 | |
| QY | 296 | -----HSDEALFGAVTEQSLPFQKHLYADGT--YQVAPLRVHCOSNLRASDVMDAIS | 346 | |
| Db | 977 | QPFALLGEAREATLQAQANQDLPFQOVLAACGGQGLFOVLFENHQOURL-----SALR | 1030 | |
| QY | 347 | SYDDRLGHLAPFGRLDIRNTGDMGSAACDFTVLLVTDGSHVNNGINGFLQQITESSHFM | 406 | |
| Db | 1031 | RLPCLLADLPWHSRE-----AKFDLQ----- | 1052 | |
| QY | 407 | PCNNRALLHCQWESSGALLVAY-YDHNVIDSLOTTRLLQOFGHLI-----KCLQSP | 457 | |
| Db | 1053 | -----LQSEEDARGTLTNFDYAADLFEASIRRFAAQYLELLFQVAEDPQRCL---- | 1101 | |
| QY | 458 | LDLSSMAENVLMTDYDRAETESNPSOLEVQDTLIHHEMLKAVSHSPKTKAIOAW--GDW | 516 | |
| Db | 1102 | -----GDIALVDAEQAAHLAENGSAPEPARAWLPPELLERQLAQSAERVALE--WDGGS | 1154 | |
| QY | 517 | TYSELNVNSRLAVHIKSLGLRAQQAIIPIYVFEKSKWVIAISMLAVLKSGNAFTLIDNPDP | 576 | |
| Db | 1155 | GYAELHARANRAHLAHLKGV--GPDVRVAICAERSPOLLVGLLAIVKAGGAYVPLDDPYD | 1213 | |
| QY | 577 | PARTAVVITRATVALTSKLHRETQVKLVGRCVVVDDELQSVASDDESSSLTKSODLA | 636 | |
| Db | 1214 | SERLAYMADSGVELLTLTQAHLEFURPAGBVGTPICLDSLUKLDNWPFSQAPGLHLHGNDLA | 1273 | |
| QY | 637 | YVIFTSGTGDPKIMEHRAFSSCALKFCASLIGNSDTRALQFGTHAFGACLEIMTTL | 696 | |
| Db | 1274 | YVIYTSGSTQCPKGVGNTHAALAEERLQWQATYTLDGDDVLQMKAPYSFDVSVWCEFWPL | 1333 | |
| QY | 697 | INGGC-----VCIPSDDDRMNIPISFINRYNVNMWMTSPYMGTSPEVDVPGLAT----- | 747 | |
| Db | 1334 | VT--GCRVLVLAAPGEHRDPARLVELRVFGVTTLHFVPLQLQFT--DEPGVACGSLRRL | 1390 | |
| QY | 748 | VLVGEOMSSSV--NAIW--APKLQLLNGYGQSESS-----ICPASNMSTBNMNGRAVGA | 799 | |
| Db | 1391 | FSGGEALPAELRNRLVRLORLPAVALHNRYGTFETAINVTHWQCRADGERSP--IGRPLGN | 1448 | |
| QY | 800 | HSWVIDPNDINRLVPIGAVGELIESPGIARDYIVPPPEKSPFTTIDIPSWPANTFPDG | 859 | |
| Db | 1449 | VVCRVLDAEFN--LLPAGVAGELCTGGGLGLARGVLGRPALSAERFVADPFS-----AAG | 1500 | |
| QY | 860 | AKLYRTGDLARYASDGSIVCLGRIDSOVKTRGORVELGATETHLRQO--MPDOLTIIVEA | 917 | |
| Db | 1501 | ERYLTGDRRWADQVLEYLIGRLDQOQKURGRIEPEELQARLLAQPGVAQVAVVIREG | 1560 | |

| | | | | | | | | |
|----|------|-----------|------------------|-------------------------------------|---------------------|---------------------|------------------|------|
| Qy | 918 | TKRSQANS | SLIAFLIGSSYFGNRP | SDAHLIDHDATAKAINIKLEQVLPHRSIPSYFICM | 977 | | | |
| Db | 1561 | VAGSQ | -----LVG-YTGA | GAEE--EAEQNR | LAALQAELPEYVPTQLMRL | 1606 | | |
| Qy | 978 | LELPRTATG | KIDRRRLKIMGKDI | LKOTQGAIVQOAP | PVPF-----ADTA | AKLH | 1028 | |
| Db | 1607 | AQMP | LGSPGKLDTRAL | -----PEVMOORE | HEVPRTELORRIA | 1644 | | |
| Qy | 1029 | SIWOSUG | IDPATVNVGATF | FELGGSNITAIKMYNMAR-SVGM | DKLVSNYIQTPLAGI- | 1086 | | |
| Db | 1645 | AIWSEV | LGL--PRVGLR | DDFELGGHSELLATRI | VSRTFQACDVPEL | PLRALFEASELEAFC | 1702 | |
| Qy | 1087 | -----SAYV | GKDPJLSTYILPK | STHEGPEQSYSQGR | LMFLDQDVGSLWY | LIPYAVMRGP | 1142 | |
| Db | 1703 | EQVRAA | QAAGRTDSHGAI | RIIDRQPVPLSY | SQQRWMLQLEP | DPSPAYNVGGLARLSG | 1762 | |
| Qy | 1143 | VNVDA | LBRALAALEQR | HETLRTTFEDODG | VGCVQIVHEKLS | EEMKVIDLCS | DLDPFE--- | 1199 |
| Db | 1763 | LDVAR | FEAALQALVQR | HETLRTTTPS | VDGVPQVRVHG | DGGLHMDWQD | FSA | 1822 |
| Qy | 1200 | --VLNO | QOTTPNLSSEAG | WRATLLRIGEDDH | ILTIVMHHITISD | GWISDVLRRDLNOLYS | 1257 | |
| Db | 1823 | QTLADS | EAHRPFDLES | GPLLRVCWVMAE | REHYLVVTLHL | VITEGWAMDIFARE | LGALYE | 1882 |
| Qy | 1258 | AALKSD | PLSALTPLPTQ | YSDFAKWQD--QFTE | QEQKQUNYWK | KOLKQDSSP-AKIP | TDF | 1314 |
| Db | 1883 | AFLLD | RESPL--PLP | VQVLDYSVWQ | REWLESGERQ | LDYKKAQLGNEH | PLLEPGDR | 1939 |
| Qy | 1315 | ARPALL | SGDAGCVHVT | DGELYQSLRAF | CNEHNTTSFV | LLAAPRAHYRLT | AVEDAVIG | 1374 |
| Db | 1940 | PRPVP | QSHQDLYRFD | LSPELAERVR | FRFNAARGIT | MTMTATLUALLY | RYSGQDURIG | 1999 |
| Qy | 1375 | TPTANR | NRPELIEDIG | CFVNTQCMKRI | INDHDTFGTL | INOVKATTTAA | FENEDIP | 1434 |
| Db | 2000 | APVAN | IRPSEGLI | AFNLQVLR | CLRDGOMSV | GELLEQVQRT | VIDGSHQDLP | 1509 |
| Qy | 1435 | SALQP | SRDLSSTPLA | OLIFAVHSQ | KDLGRFKFOG | LESVPVPSKAY-----TRF | DMOE | 1487 |
| Db | 2060 | EALQP-PR | SAAYNPLFQ | VMCNVQ-----RWE | FQOTRLQAG | MTVEYIANDARAT | KFDLNL | 2112 |
| Qy | 1488 | HLFQET | SLKGSVNA | FELEFKMET | VENVRVFFELL | RNGLOSSRTP | VSILPLDTG | 1547 |
| Db | 2113 | EVTDLQ | RGLCCLYSR | DLDEPR | IARMAGHWQ | NLLLEALLGD | PPORRTAELP--FAAE | 2169 |
| Qy | 1548 | KDLVN | KHVDPRESS | LAD-----VFOT | VSAYPDSLAV | VDVSCRLTYVTE | LDROS | 1603 |
| Db | 2170 | RKQLLAG | TAG--EAG | LQTLHCLFA | ARVNAASQ | APALTFAGOT | LSYAE | 2226 |
| Qy | 1604 | WLRRS | MPAETLVAF | APRSCETTIV | AFGLKANLAY | PLDVRSPS | ARVQDIL--SGL | 1661 |
| Db | 2227 | VLRSH | GVGPEVRV | GALERS | LEMVVGLLAIL | KAGGAYVPL | DPPEPLER | 1712 |
| Qy | 1662 | PTIVL | IHDHTAP | PDIEVTNVE | FFVIRI | DAINDSNA-----DG | FEVIEHDS | 1766 |
| Db | 2286 | ---LLSH- | -----AAL | FALGELPAG | VARWCLEED | GPALDAED | DPAPLAALS | 2328 |
| Qy | 1713 | ---LAV | VLTSSTGRPK | GMIEHRVIR--TVT | SGICPNYP | SPSETRMAH | MATIAFDGAS | 1766 |
| Db | 2329 | GPQO | QAVLIYTS | SGTGPKG | VANVSHGEL | AMHCAVIECF-GM | RAEDCEL | 1766 |
| Qy | 1767 | YEI | SALLFG | RTL------CYD | YMTTLDAR | AKDVFFREH | YNA--ASH | 1816 |
| Db | 2388 | ERLLA | PLLCGARV | VLRAQOG | WGAEEICEL | IRAE | GVSLGFT | 1816 |
| Qy | 1817 | LRVPR | RLSRTL | MMFFLV | VTDS | TADALDAQ | GLYOG---VQ | 1816 |
| Db | 2448 | VR----- | MCITGE | ALTGEHLQ | RIRQAF | ASPFPNAY | GTETV | 1816 |
| Qy | 1872 | IDT | ESFING | VPIGRAL | NNSGAYVVD | PEQQLVG | IVGM | 1816 |
| Db | 2495 | --ER | LEE | GAASVPI | GSVGVAR | VAYIID | ADLALV | 1816 |
| Qy | 1931 | RFVH | ITVNDOT | VKAYRTG | DRVRYRIG | DGLIEFFGR | MDTQFKIR | 1816 |

| | | | |
|----|------|---|------|
| Db | 2554 | RFVDPFPAEGGRUYRTGDLVRL-CDNGQVEYVGRIDHQVKIRGFRLEIEIEARLLEHP | 2612 |
| Qy | 1991 | SVRDAAVYLOQNEQDAPEIILGFV---VADHDHSENDKQGSANQVEGWQDHFSGMYSDIG | 2047 |
| Db | 2613 | QVRE-ALVVALDSPGKOLAGYVASAEQD---EDAQAA- | 2648 |
| Qy | 2048 | EIDPSTTIGSDFKGWTSMYDQSQDFDEMHEWLGETTFTLHDNRSLGNVLBIGTSGMILF | 2107 |
| Db | 2649 | -----LREAKT----- | 2655 |
| Qy | 2108 | NLDSRLSEYVGLPSPSAAAFVNKATESIPSLA-GRAKVGVGTATDIGQVDDLHPDLVL | 2166 |
| Db | 2656 | HLKQQLPDY--MVAHLL-----LLASLPLTANGKLDRLRALPAD-----PAL | 2696 |
| Qy | 2167 | NSVIQYPPSSEYLAETIADTLIHLPNVQRIFFGD-----VRSQATNEHF- | 2209 |
| Db | 2697 | NRQAYEAPRVLBEQLAGVYREVLNVBERVGLGNFFELGSDSILISQVSRARQLGIHFS | 2756 |
| Qy | 2210 | -----LAARAHTLGNKATKDDYR-----QKMAELEDMEERELL | 2243 |
| Db | 2757 | PRDLFQHTQVQSLLAAVAARHSQASQAEQPVQGSALTPIOHWFFDPLPLARREHWNQALL | 2816 |
| Qy | 2244 | EPAFFTSKDRFPGLVEHVEILPKNME-AVNELSAYRYAA-----VVHVRGSLGDVY | 2295 |
| Db | 2817 | QPROAIDL-----GLLRKSLQRLVEQHDALRALAFQVDGEWLAQHRPLREOELL | 2865 |
| Qy | 2296 | --LPVEK-DWDIDFOANQLOKSL---GDLKSSDAAIMAVSKIPEITAFERQVVASL | 2348 |
| Db | 2866 | WHVPQSFDECAELFAKA--ORSULDEQGPLLRA-----VLVDGPAGQORLLAI | 2913 |
| Qy | 2349 | NSNIDE---WOL-----STIRSAEGDSLSVPDIPRIAGEAGFRVEVSARQWSONGA | 2399 |
| Db | 2914 | HHLVVDGVSNRVLLEDLQQVYVQFAEG-----AEPALPAKTSAFRDWA--GR | 2958 |
| Qy | 2400 | LDVFFHCCSQ-----GRTLNFPTDHHLRGSD-----LLTNRPLO | 2435 |
| Db | 2959 | LQA---YAGESUREELGWQARLGGQPVWPCDRP-QGDNRREALAESVSLRDLDPQTRQ | 3014 |
| Qy | 2436 | RLQNRRIA--IEVRERLRSLLPSYMI-----PSNIVVLD----- | 2467 |
| Db | 3015 | LLQQAAPAYRTQVNDLLLTALARVLCWSPSTVLQLEHGREALFDDIDLTRSVGWFT | 3074 |
| Qy | 2468 | -----KMPUNANGKVDKRLSRAKVPK-----QOTAAPLPTTPISE | 2505 |
| Db | 3075 | SAYPLRLTPAQSPGE-SIKAIKEQLRAVPHKGLGYGLRYLADPAVRQMAALPTAPIT- | 3132 |
| Qy | 2506 | VEVILCEEATEVFGMKVDITDHFENLGGHSLATKLISRIDQRLKVRITVKVDVEDHPVFA | 2565 |
| Db | 3133 | -----FNYLGO-----FDQS-FA | 3144 |
| Qy | 2566 | DLASVIRQGLGLQPVSDGQDQRSAHMAPRTEATILCDEFKVLQFQVIGITDNFFDLG | 2625 |
| Db | 3145 | D-----ALFQILDOPTG-----PIHDEQAPLPNELS--VDGQV-----YG | 3177 |
| Qy | 2626 | GHSMLATKLAVRIGHRLDITTVSVKVDVDPHVLQLAIA-----LDNLVQSKTNEIVGGREM | 2681 |
| Db | 3178 | GELV-----LRWTYS-RERYDARTVNELAQAYLAELQALIEHCELDGAGGL-- | 3222 |
| Qy | 2682 | AEYSPFOLLFTEDPEEFMASEI-KPOLEQEI-----IODIYPTOMKAFLEPDHTTARPR | 2736 |
| Db | 3223 | -----TPSDFPLAQLSOAQDLADALVAPAGEIEDVYPLTPMQEGLLL-HTLLEPG | 3269 |
| Qy | 2737 | PFV-----PFYIDPSTSEPDAAGLKACESLVNHLIDIFRTVFAEASGE-LYQVVLSCLD | 2790 |
| Db | 3270 | TGIYVMQDRIYDPSLDPERFAA-----AMQAVVARHEALRASPVWAGETMLQVHKPGR | 3325 |
| Qy | 2791 | LPIQVLETEDNINTATNEFLDEFKPEVRLGHPILBFTIIKQPKSMRVIMRISHALYDGL | 2850 |
| Db | 3326 | TRIEFLDWSLDPDGHEERLQALUHKRERAG-----FDLLEQDPFHRLIRLIRGEARYWFM | 3380 |
| Qy | 2851 | SLEHVVRKRLHMYNG--RSLL-----PPHOFSRMYQTA---D | 2883 |

CC gene cluster comprising the ORFs is useful as a substrate for
 CC bioengineering of antibiotic structures. An isolated polypeptide or its
 CC encoding nucleic acid sequence is useful for generating derivatives of
 CC ramoplanin, for improving production or for producing variants of other
 CC antibiotics of the peptide class. The isolated polypeptides are useful
 CC for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain
 CC in conjunction with other peptide synthetase modules and allowing the
 CC incorporation of Thr into a peptide antibiotic precursor, for modifying
 CC fatty acid structure and/or enhancing fatty acid incorporation into the
 CC peptide antibiotic structure, for production of an hydroxyphenylglycine
 CC (HPG)-containing peptide antibiotic, for enhancing secretion of
 CC ramoplanin or its variants and derivatives, for enhancing uptake of
 CC precursors for ramoplanin biosynthesis, for enhancing production of
 CC ramoplanin products or its variants or derivatives, to chlorinate HPG of
 CC a peptide antibiotic precursor, and for designing specific nucleotide
 CC probes and primers for identifying and isolating putative lipopeptide
 CC -producing microorganisms. This sequence represents one of the ORF
 CC proteins of the ramoplanin producing Actinoplanes sp. microorganism of
 CC the invention.

XX Sequence : 4999 AA;

Query Match 10.1%; Score 1632; DB 23; Length 4999;
 Best Local Similarity 22.5%; Pred. No. 5.9e-121;
 Matches 738; Conservative 417; Mismatches 1123; Indels 1000; Gaps 109;

QY 43 SSRIEAIKPTCFQDMDICNAL-DKQSAIGHAVYDVPTDID-----ISRPAIAWKEIV 95
 DB 1104 AANIADVYPLAPLQEGIEFFHMMADRDSA--DVVYTPVVEFSDRLDGLAALQQVV 1160

QY 96 NOTPALFAFTSDSGKTSQVILKDSFVSWCMWSSSSSPDEVRDEAAAAAGSPRCNRF 155
 DB 1161 DRT-----DVRTSV-----WQGLPEVQVVRHARLPV-----DEV 1193

QY 156 VLLEDW---QTKKQQLVYTESHALVDVTFQORVLSRVFAAYKHEKDHRTPESSDAT 211
 DB 1194 VLRDDLDPEQNALGTAW-----MDLS-EAPLQAVVAA-----RPGDPQ----- 1233

QY 212 DTDOSVSVSSMCDNAVSAFHWQTH-----LNDLNASVFPHLSDRLMVPNPTTAE 265
 DB 1234 ----RWLAFLRH-----HLVQDHTALDILLEEAAYLAGRGD--LPEVPVPRE 1277

QY 266 HRITPLSQALNSAICRTALSILLSRYTHSDEALFGAVTQSLPF----- 312
 DB 1278 F-----VAHRLGVPREEHRYFAGLLGDVETETAPYGLLDVHSGGLASA 1322

QY 313 DKHVLADGTY--QTVAPLRVHCQS-----NLRASDVMDAISYDDRLGLHLPFLGRDIRN 365
 DB 1323 QAHRLDGLPLGRVRAFAREHGVSPATFLHLAWARVLGTLAGRDDVYFTGTVLFGRM---- 1378

QY 366 TGDNGSAACDFTQVLLVTD-----GSHVNNINGFLQOITPESHFMPCNNRALLHCQ 418
 DB 1379 ----NSGAGADRVPLGFLINTLPVRVLGAPVGDALDGLDQLTE-----LIAH-- 1422

QY 419 MESSGALLVAYDHNVID-----SLOTTRLQO----- 445
 DB 1423 ---EHAPLAVAOAANLFGORPLFTSIFNYRYARGAEPAGAALDGIIRLLSARDLTNYPLAVA 1480

QY 446 -----QFGLIKLQSPDLSS-----MAEVNLMTEYDRAEI 477
 DB 1481 VDAEGDTSLTVDAVAPADPVQVGGELLYTALRNLTAEANPGTFLAAGVGLGEDELSRV 1540

QY 478 ES-WNSQPLEVODTLIIHMLKAVSHSPTKTAIOAWGDWMTYSELDNVSRLAVHIKSLG 536
 DB 1541 VSGWNTARRVQASVPELFAERVAAPGAPAAVAGDLRWTVYADLDARSALARSILVA 1600

QY 537 LRAQAAILPVVEKSKWVIASMLAVLKSNAFTLIDPNDPPARTAQVVTQTRATVALTSK 596
 DB 1601 VTAEPSVW--VALERSADVLTAFLAVAKAGGVFPVDLSWPRARVDVATADCAAWTAVADR 1659

QY 597 LHRETQVQLGVRCVVDDELLQSVASDDFSLSLTKSQ---DLAYVIFTSGSGDPKGIWI 653
 DB 1660 -----PMTGLTVVPANR-----AGDPVALPPRPLPGAANYRMVTSYGRPKGVVT 1706

QY 654 EHR-----AFSSCALKFGASLINGSSTRALQFGTHAFGACLEIMTTLINGCVCI-PSD 707
 DB 1707 THQNVVDLYDRC---WGPT-----PRVLFHAFHAFDASSFELWVPLLTGTGVVVAPE 1757

QY 708 DDRMSIPSFINRYNNVNMMAFSPYMGTFSS--PEDVPGLATLVLVEQOM-SSSVNAIW-- 762
 DB 1758 SDTDGVRLQIRAHLELTHVHTAGLLRVLAEDPSCFAGLTEVLTDGDDVVPAEAVRRVLD 1817

QY 763 APLKOLLNGYGOSESSICFASNMSTEPN---NMGRV-CAHSWVTDNDINRLNPIGAV 818
 DB 1818 NPGVVRQLYGPTE-VTLCATQHVVRPEPSVLPVIGRPLDNTRVYVLD--GLLQPVGVGT 1874

QY 819 GELVIESPGIARDYIVPPPEKSPFTDIPSWYPANTFPDGAKLYRTGLDARYASDGSIV 878
 DB 1875 GELYTAGAGVARGYADMPCTTAERFVAD-----PFTAGGLRYTGLDVRMTGEGELY 1926

QY 879 CLGRIDSQVKIRQORVELGAIETHL-----RQMPDDLTIVVEATKKSQSAN 925
 DB 1927 FAGRADDOVKIRGYRVEFEVEAVLAALPGVSQAIVIVREDVPGD-----KR----- 1973

QY 926 STSLIAFLIGSSYFGNRPDSDAHILDHDATAKINIKLEQVLPKHSIPSFYICMLELPTAT 985
 DB 1974 ---LVAYLVAAP-----ETVEAARAHAEQRLPSYLVPSAFVQDLDALPLTN 2016

QY 986 GKIDRRRLRMGKDILDKOTOGAIVQOAPAPIPVFADTAALKHSTWVQSLG-----IDP 1039
 DB 2017 QKVDRAAL-----PAPLGEAGAGRAPADAREELVGAFAEVLDL 2056

QY 1040 ATVNVGATTFELGGNSITAIKVMNMARSVMGLKVSNIYQHFTLAGISAVVVGDPUSYTL 1099
 DB 2057 GRVGDDDFPFGALGHSLALALVERLRROGLGVSRAVEDARTPAALTFRGGDADRPA 2116

QY 1100 IPKSTHEGVEQSYSGRLWFLDQVLSLWTLIPYAVRMGRPVNVVDALRRALALEORH 1159
 DB 2117 LRAGARPARKPLSYAQRRUNFLAQLEGPSATTNIPVALRLEGLDORDALTAALRDVVARH 2176

QY 1160 ETLRTTFEDQDGVQIVHEKLESE--MKVIDLCGSDLDPPFVLEQEQTPFNLSSEAGM 1217
 DB 2177 EVLRTVFTVADGEPQWHLDPARAEPALPVVDVPGARVE--EAVAEAAAAYFDLAREIPL 2234

QY 1218 RATLRLGEDDHILIVMHIIISDGSIDVLRDLNQLYSAAKSKDPLSALTPLPIQY 1277
 DB 2235 RAVLLAPDGTHTVLVLVHLHIAADGWSMRPLARDLATAY-AAARRRQAPES--ETLPVQY 2291

QY 1278 SDFAKWOKDOFTEQ-----EKOLNWKOLKDDSSPAK--IPTDFARALLSGDAGCVH 1328
 DB 2292 ADYALWQRDLLOSDSDSPASLSRQTAHWRERL-DGVPEELDLPADRPRAAASHRGHLHS 2350

QY 1329 VTIDGELXQSLRAFACNEHNTTSFVLLAAAFRAAHVRLTAVEDAVIGTPIANRNRPELEDI 1388
 DB 2351 AEIPADVHRSRLRRVAADHGCATVFMTLQAAVAVLLSRLGAGTDVPIGTVVAGRADALENL 2410

QY 1389 ICGFVNTQCMRINIDHHDFTGLINQKATTTAAAFENEDIPEERVVSALQPGSRDLSSTP 1448
 DB 2411 VGGFVNTLVRLTDLGDPRLTDVGLQVRELTLRALAHQDVPPEKLVLEETP-ARSLARHP 2469

QY 1449 LAQLIFAVHSQRDLGRFKFGQLESVPVPSKAY--TRFDME-----FHLFEQETDSLKGSVN 1501
 DB 2470 LFQVWVTLDDGGPDGA-ELPGLAMSVVPTGAVPAKFDLDLTETETFDAGEPAGLRVDLI 2528

QY 1502 FADELFKMTVENNVVRFVEILRNGLQSSRTFVSLPLPLTDGIVTLEKLDVNVKHHVDYPR 1561
 DB 2529 AAADLFDAGTAARLAGLYLSRVL--GVLA--DPRRRRLAEVDPLEAEESRLMLAAGEBPAPA 2585

QY 1562 ---ESSLADVFOFQVSAYPDSLAVDSSCRLTYTTELDRQSDILAGWLRRRSMFAETLVAVF 1619
 DB 2586 LPEITVAALVAPQOCARTPGAVAVTGPASLTLYAELDRAARTIWLRRHAGGAPGAACVVL 2645

QY 1620 APRSCETIVAFVGVKLANLAYLPLDVRSPSARVQDILSLSGPTIVLIGHDTAPPDIEVT 1679
 DB 2646 MERSAELVAVLLGVMRAGAAYVPVDPAYPAERIRFVVVTDARACVV---SESASAGLVPD 2702

QY 1680 NVEFVRIRDALNSNADGFEVIE-----HDSKPSATSLAYLYTSGSTGRPKGVMEHR 1734
Db 2703 GVPCLAI-----DPAAPAAEPAPGDDPGDAAGPRDDPAYIITSGSTGTPKGVVSHR 2757
QY 1735 VIIRTVT-SGCIINYPSETRMAHMTAFDASVEIYSALLFGRTLVCDVMTTLDAARL 1793
Db 2758 NVALLATPLPGFAGDEWFSHFSAFDFSVWELMGALTHGRVVVVVYVRSRPRDF 2817
QY 1794 KDFVFRHVAASHVTSOSSDVLPRV---PRRLSRTLMFFFLVTVTSTAPDADAOGLYQ 1850
Db 2818 WELVREGVTLSQTPSAFAQLMAAAGDDDRDALRFVVF-----GGEALD-PCRLA 2867
QY 1851 G-----VOCYNGVGPTEGWMSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEQOL 1902
Db 2868 GWLARRPDKRLVNMVIGITETTVHTTYQHIAPTT---GSVIGRGLPGFLGYLDALRP 2924
QY 1903 VGIYMGELVVTGDGLARGYSK-ALDENRFVHTVNDQVKAIRTGDRVRYRIGDGLIE 1961
Db 2925 VPAGVGEVYARGPQVARGVIGRPLTAERFV-ASFPAGGERMYRTGDVARW-TADGRLV 2982
QY 1962 FFGMDPTQKIRGNRIESAIEAALLRDSVSRDAAVVLOQNEQOAPAILFVYVADHDSE 2021
Db 2983 FAGRSDDQIKIRGFRIEPEGEVEAVLAAGPGVSOAAVIVR-----3022
QY 2022 NDKGOSANQVEGWQDFESGMYSDIGEIDPSTTGSDFKGTWSYDGSQIDFDEMHEWLGE 2081
Db 3023 -----DVPG-----3026
QY 2082 TTRTLHNRSLGNVLEIGTSGMILNLSRLESYVGLPESRSAAAFVNKATESIPSLAG 2141
Db 3027 -----DKRLVAYV-----3034
QY 2142 KAKVQVGTATDIGQVDDLHPDLVLVNSVIQYFSPSEYLAETDLIHLPNVQRIFFGDVR 2201
Db 3035 -----VG-----GDA-3039
QY 2202 SQATNEHFLAARAIHTLGNKATKDDVRQKMAELEDMEELLVEPAFTSLKRFPLGVEH 2261
Db 3040 -----3039
QY 2262 VEILPKMEAVNELSAYRYAAVVHVRGSLGDELVLPEKDDWIDFQANQLNOKSLGLDLK 2321
Db 3040 -----3039
QY 2322 SSDAIAVMSKIPFEITAFERQVVASLNSIDEWQLSTRSSAEGDSSLVSPDIFRIAGE 2381
Db 3040 -----ETLRSQAQ-----3047
QY 2382 AGFREVSSARQWSQALDVAHFHCCSQGRTLNVNFTDHLHRLGSDLLTNRLQRLQNR 2441
Db 3048 -----VG-----OR-----3049
QY 2442 IAEVRELRKSLPYSMIPSNIVLDKMPNLNANGKVDKRLSKRSRAKVPKQQAAPLPTF 2501
Db 3050 -----LPGVLVSFAFVELDLPLTVNGKLDLR-----ALPVPDY 3083
QY 2502 PISEVEILCEATEVFGMKVDITDHFHNLGGHSLATKLISRIDQRLKRVITKDVDFH 2561
Db 3084 -----3083
QY 2562 PVFADLASVTRQGLGLQPVSDGQODRSAAHMAPRTETAILCDEPAKVLGFO-VGIDTN 2620
Db 3084 -----GRDAGGGRAPANAREEVLGRAFAEVLGVERGVEDD 3119
QY 2621 FFDLGGHSLMATKLAVRIGHRLDTTVSKVDVDFHPVLQALATLDNLVQSKTNEIVGGRE 2680
Db 3120 FFALGGHSLVLSVLERL-RRQISVVPVRLFTTPTPAGLA-----EAVG--D 3164
QY 2681 MAEYSPQLFTEDPEEFWASEIKPOL-----ELQEI-----IIDIYPSQ 2721
Db 3165 GAVVVPNNLI-----PED--AAELTPEMLPLADLTADLAVVAVSVPGAANTADVPLAP 3218
QY 2722 MQKAFLEDHTTA-RPRFV---PFYIDFPSTSEPDAAGLIKACESLVNHLIDFRT-VFAE 2776

Db 3219 LQEGIFFHHMADRDSADVVYVTPVTVFEFDSRDLR--GFLAALQVVDRTDVTYSVMQ 3276
QY 2777 ASGELQVVLSCLDLPIQVITETEDNTATNEFLDEFAKEPVLGH-PLIRFTIUKTKS 2835
Db 3277 GLREPQVQVWRHARLPIDEVELHEGTDPA--EQIALATERVDLDRAPLIRTTAAVPGS 3334
QY 2836 MR--VIMRISHALYDGLSLHVVYRKLHMLYNGR--SLPPHQSRYVMQYTAGD--RESGH 2889
Db 3335 GRWALLRIHHLVQDHTTLDVLLGELRAFLEGRGDELPEVPFREFVAQARLGVPREHE 3394
QY 2890 GFWRDVION--TPMTILSDDTVVGDNDATKALHLSKIV--NIPSOVLRGSSNI-ITOATV 2945
Db 3395 RYFAELGLGVDTETAPYGLTEVHGDSA--AVHSRREVDDDLAARLHRLARSLSVSPAL 3452
QY 2946 FNAACALVLSRESDSKDVFGVGRVSGRGLPVEYQDIVGCTNAPVVRRAHIESSDYNQLL 3005
Db 3453 FHLARVLCANSGRDDVFGTVLFGRMNSGAADRVOGLFINTVLPVVRVLAAGSTRDAL 3512
QY 3006 HDIQDQY--LLSLPHETIGFSDLKRNCTDMPEAITNFSCCITYHNFEYHPESOFEOQORVE 3053
Db 3513 TGLRDQLAGLLVHEHAPLALAQRAAGITDGSPLFASI-----FNRYHNODDPAASAGLE 3566
QY 3064 MGVLTFVNIEMDEPLYDLAAGEVEPDGAGLKVTVIA 3101
Db 3567 -GIRTVYSAHTNYPL-DASI--DVTGDRFAITVNAVA 3600

RESULT 9
AAR34712.
ID AAR34712 standard; Protein: 3588 AA.
XX
AC AAR34712;
XX
DT 17-AUG-1993 (first entry)
XX
DE Bacillus subtilis srfa operon ORF1 prod.
XX
KW Multienzyme complex; surfactin synthetase; MCSS; ORF; surfactant.
XX
OS Bacillus subtilis.
XX
FH Key Location/Qualifiers
FT Peptide 139..145
FT /note= "repeat sequence"
FT Peptide 1184..1190
FT /note= "repeat sequence"
FT Peptide 2223..2229
FT /note= "repeat sequence"
FT Peptide 3255..3261
FT /note= "repeat sequence"
FT Peptide 606..616
FT /note= "repeat sequence"
FT Peptide 1653..1663
FT /note= "repeat sequence"
FT Peptide 2698..2709
FT /note= "repeat sequence"
FT Peptide 858..880
FT /note= "repeat sequence"
FT Peptide 1897..1919
FT /note= "repeat sequence"
FT Peptide 2930..2949
FT /note= "repeat sequence"
XX EP540074-A.
XX
XX
PD 05-MAY-1993.
XX
PF 03-OCT-1992; 92EP-0203037.
XX
PR 09-OCT-1991; 91IT-OMI2683.
PR 02-SEP-1992; 92IT-OMI2044.
XX

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------|-------|-------|-------------|----------|-------|-----|-------|-------|------|-----|-----|-------|-------|----|----|---|---|---|---|---|---|---|---|---|---|---|-------|----|---|-------|----|---|-------|----|------|---|-------|-------|----|---|----|---|---|-------|----|-------|----|---|------|---|------|------|---|------|------|------|
| 2664 | ---- | ----- | DGSRILSLPLD | NDEENPEN | PETA | VA | TENLA | MYIT | STCT | QOP | KGV | VVEHH | 2715 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1735 | VI | IRT | VTSG | CIPNY | ----- | P | SETR | MA | HMA | TI | AF | DG | SY | EI | YS | A | L | L | F | G | R | T | L | V | C | V | D | Y | M | T | T | L | 1789 | | | | | | | | | | | | | | | | | | | | | | | |
| 2716 | AL | VNL | ---- | ---- | ---- | CF | WH | D | A | F | S | M | T | A | E | R | S | A | K | A | Y | G | F | D | A | S | I | W | E | F | T | S | I | G | A | E | L | H | V | E | E | A | R | L | D | 2791 | | | | | | | | | | |
| 1790 | AR | AK | DV | F----- | ----- | PRE | VNA | A | SH | V | T | S | S | Q | D | V | L | R | V | P | R | L | S | T | R | T | L | M | F | F | F | L | 1832 | | | | | | | | | | | | | | | | | | | | | | | |
| 2772 | IV | R | ND | Y | F | E | T | G | Y | T | I | T | F | L | P | T | Q | L | A | E | Q | F | M | E | L | E | N | T | S | L | R | V | L | L | T | G | D | ----- | KL | K | AV | K | P | Y | T | 2824 | | | | | | | | | | |
| 1833 | V | T | D | S | T | A | P | D | A | L | D | A | O | G | L | Y | Q | V | Q | C | Y | G | P | T | E | N | G | M | S | T | I | P | D | S | T | E | S | F | I | N | G | V | P | C | I | R | N | L | N | S | G | 1892 | | | | |
| 2825 | L | V | ----- | ----- | ----- | NN | G | P | T | E | N | T | V | V | A | T | S | A | E | L | H | P | E | G | S | L | ----- | IG | R | A | T | A | N | T | 2860 | | | | | | | | | | | | | | | | | | | | | |
| 1893 | A | Y | V | D | P | E | Q | L | G | I | C | V | M | G | E | L | V | T | G | D | L | A | R | G | Y | S | K | D | A | L | D | E | ----- | NR | F | V | H | I | T | V | N | D | O | T | V | ----- | RA | Y | R | T | 1947 | | | | | |
| 2861 | V | I | L | E | G | N | Q | V | P | E | G | V | A | G | E | L | C | V | A | G | L | A | R | G | Y | L | N | R | E | ----- | DE | T | A | K | R | F | ----- | VA | D | F | V | P | C | E | R | M | Y | T | 2915 | | | | | | | |
| 1948 | G | D | R | V | R | I | R | I | G | L | L | E | F | F | G | R | M | T | O | P | K | I | R | N | R | I | E | S | A | E | I | A | A | L | L | R | D | S | V | R | D | A | A | V | V | L | O | O | N | E | D | A | P | 2007 | | |
| 2916 | G | D | L | V | K | W | T | ----- | GG | E | I | E | I | G | R | I | D | O | Q | V | K | R | G | R | I | E | L | S | E | I | E | V | O | L | A | E | S | E | V | O | D | A | A | V | A | T | V | K | D | K | G | G | N | T | 2973 | |
| 2008 | E | I | L | F | V | V | A | D | H | D | H | E | N | K | G | S | A | N | O | V | E | G | Q | D | H | F | E | S | M | Y | S | D | I | G | E | I | D | P | S | T | I | G | S | ----- | DF | 2058 | | | | | | | | | | |
| 2974 | A | I | A | Y | T | P | ----- | S | A | D | I | E | A | L | K | S | A | K | E | T | L | P | D | I | M | I | P | A | F | W | T | L | M | E | L | P | T | A | N | G | K | V | D | R | K | A | L | E | P | D | I | 3030 | | | | |
| 2059 | K | G | M | T | S | D | G | S | I | D | F | E | M | H | E | M | L | G | T | T | R | L | H | D | N | R | L | S | G | N | V | L | E | I | G | T | S | G | M | I | F | L | N | D | S | R | L | E | S | Y | -- | 2116 | | | | |
| 3031 | E | A | G | S | G | E | Y | K | A | P | T | T | O | E | L | L | A | G | I | W | D | V | L | C | M | S | E | V | G | T | D | N | F | F | S | L | G | ----- | G | D | S | I | K | I | O | M | A | S | R | L | N | O | H | G | W | 3089 |
| 2117 | ----- | VG | L | P | S | R | S | A | A | F | N | K | A | T | E | S | I | P | S | L | A | K | A | K | V | Q | V | G | T | A | T | D | I | G | V | D | D | L | H | P | D | L | V | L | N | S | V | I | 2170 | | | | | | | |
| 3090 | K | L | E | M | K | O | L | N | O | H | P | T | ----- | TE | L | P | Y | E | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10

AAB83971
ID AAB83971 standard; Protein; 1537 AA.

XX AAB83971;

XX
DT 06-AUG-2001 (first entry)

xx DE Amino acid sequence of a type I polyketide synthase.

XX Metabolic pathway operon; polyketide antibiotic;
KW type I polyketide synthase.

XX Unidentified.

XX
PN WO200140497-A2.

XX 07-JUN-2001.
PD

| | | |
|----|---|--|
| XX | 27-NOV-2000; 2000WO-FR03311. | |
| PF | | |
| XX | 29-NOV-1999; 99FR-0015032. | |
| PR | | |
| PR | 07-JUN-2000; 2000US-0209800. | |
| XX | | |
| XX | (AVET) AVENTIS PHARMA SA. | |
| PA | | |
| XX | | |
| PI | Jeannin P, Pernodet J, Guérineau M, Simonet P, Courtols S; | |
| PI | Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K; | |
| PI | Frostegard A; | |
| XX | | |
| XX | WPI: 2001-374849/39. | |
| DR | | |
| XX | N-PSDB; AAF90034. | |
| DR | | |
| XX | | |
| PT | Collection of nucleic acids from environmental samples, useful for | |
| PT | identifying e.g. genes encoding polyketide synthases and derived | |
| PT | antibiotics | |
| XX | | |
| PS | Claim 78; Page 313-318; 356pp; French. | |
| XX | | |
| CC | The specification describes a method for the preparation of a collection | |
| CC | of nucleic acids from organisms in a soil sample. The method comprises | |
| CC | milling a dried sample to produce microparticles; suspending these in | |
| CC | liquid buffer; extraction of nucleic acids from the microparticle; | |
| CC | passing nucleic acid-containing solution through a molecular sieve; | |
| CC | passing nucleic acid-enriched fractions through an anion exchange | |
| CC | chromatography material; and recovering fractions containing purified | |
| CC | nucleic acids. The nucleic acids are sources for sequences that encode | |
| CC | either operons involved in a metabolic pathway (specifically polyketide | |
| CC | synthesis) or polypeptides, particularly for production of therapeutic | |
| CC | or agricultural compounds, especially polyketide antibiotics. AAB83971-76 | |
| CC | represent type I polyketide synthases encoded by cosmid a26q1. | |

| | | |
|----|----------|----------|
| XX | Sequence | 1537 AA; |
| SQ | | |

```

Query Match          9.7%; Score 1572; DB 22; Length 1537;
Best Local Similarity 29.3%; Pred. No. 4.8e-117;
Matches 476; Conservative 269; Mismatches 670; Indels 208; Gaps 51;

QY      1040 ATYNGATFELGGNSITAIKMYNMARSVGMDLKV---SNIYQHT-----LAG 1085
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      6 SAVDLQOPLVRMGDLSLMAVOLRN--RIDTDLRVLLPMVRFLDGFSVAELARDLSDLG 62

QY      1086 IS--AVVKGDP LSYTLIPKSTHEPGEVQSYSGRGLWFLQDLQDVGSLWYLIPIYAVRMGPV 1143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      63 LSRTTVAPEAAQASVPALSY--PL--SAGQOALWFIYRSAPESPAYNTAWTARAGAF 118

QY      1144 NVDALRRALAAEQRHETLRTTFEDQDGVGVQVJLHEKLSSEMKVIDLCGSDLDPPFVLNQ 1203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119 DPQALRRSQDLVDRPALKTTTAEAGGAPQVQTVSHSVDPDFEIP--CSPD--DEAVLIDG 176

QY      1204 EQTPTFNLSSEAGWRATLRLGEDDHTLITVMHHIISDGWSIDVLRDLRNQLYSAAALKDS 1263
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      177 VFHAPENL--GENCFSRLLVQSKDQGVLAIVVHHILLDFWSLLVMVDLRSIYLA--RTA 233

QY      1264 KDLPSALTLPLPIQYDFAKWKQKQDF--IEQEKOLNTWKQKLDSSPA--KIPTDFARPALL 1320
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      234 GGP--VPAPVSAFAFVRWQNELLAGTEGERLWNYWSSQSGQLVNLNPSDRPSPVPQ 291

QY      1321 SGDAGCVHVTIDELYQSLRFAFCNEHTTSFVVLAAAFRAAHRYRLTAVEADVGTPTANR 1380
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      292 SFRGNSHSFRIEPAITAKLALARRQNATLHATLMAAFQVLLSRWTSQEELTGTNGR 351

QY      1381 NRPELEDIIGCFVNTQCMRINIDHDTFGPLINQVKATTTAAFENEDIPFVRVVALQPG 1440
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      352 TQPEFADLVGYFVNPVILRGELSGDPPDFNTVLARIQTLILGALHEQEPYVARIVERLGP 411

QY      1441 SRDLSSTPLAQLFAVHSQDL-----GREKFGQ--LESVPVPSKAYTRFDM 1486
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      412 LR-----VLFVLOQPHRIPESVFMQLGSGGRMAGWSLTLESAMP LR--QSRFLLD 461

QY      1487 FHLFQETDLSLKGSVNFADELFKMETVENVYRVEFFELRNGLQSSRTPVSLPTLDGIVTL 1546

```


Best Local Similarity 25.9%; Pred. No. 1.8e-103;
Matches 496; Conservative 320; Mismatches 812; Indels 287; Gaps 64;

QY 264 AEHRITPLSOKALSNAICRTAL-----SILLSRYTHSDEALFGA----- 304
Db 1122 AEHRI-----KTLSDVPRTAAIKDCWAGHRVTLHSLVOPAMHKLFAIGSGPVTWGT 1175
QY 305 -VTEQSLPEKHYLADGTQYQVAPLRVHCOSNLRASDVMAISYDDRLGLHAPFGLRDI 363
Db 1176 VISGRGLPVGIEDSVGLYINTLPSIVD-HAGQAGQTIAEALADIQESINRMNSRSAAVE 1234
QY 364 RNTGDNGSAACDFTVLVTGDSHVNGINGFLLQITESSHFMCNNRALLH---COME 420
Db 1235 FMVOTGSKRRLPETLL-----VLENYPILLSSEEAALHHEHLCFEK 1276
QY 421 SSGALLVAY-----VDHNVIDSLOTTRLLQOFGHLIKCLOSP 457
Db 1277 SYSDSKVDYFVAVVAREDDGRFEINLWYAGELEFDGTVDNLLGT-LDALFQQAADFDAP 1335
QY 458 LDLSMAEAVNLMTEDYDRAETESWNSQPLE-VODTLIHHEMLKAVSHSPTKTAIQAWGDW 516
Db 1336 VE-----RLELVCAADRRRDEWNRTEFEARTKTHAGFEIAERWPDIAVVGSENRL 1390
QY 517 TYSELDNVSRRLAVHKSGLRAQQAALIPYVFEKSKWVIASMLAVLKSGNAFTLIDPNDP 576
Db 1391 SYRELNERANKIAAFLAARVAPKDEFGVLWEKSEWMAIILGVWKAGAAYVPIDESY 1450
QY 577 PARTAQVVTOTRATVALTSKLHRETQKLVG--RCVV--VDDELQVSASDDFSSLTKS 632
Db 1451 DDIRIFEDTQARLVLTDRAYARLDRVADGPFCEVFGVDLPLEPYPAANPRSAAT-S 1509
QY 633 QDLAYVFTSGTGDPKGIMIEHRA-----FSSCALKFGASLGINSIDTRALQFGTHAFGAC 688
Db 1510 TDLAYALYTSGTGRKAVLIERGVVNLHTSLERLFDLSRD-RGDAVLSFSNYVDFHF 1568
QY 689 LLEIMTTLINGCVCIPSDDRMNS--IPSFIRNRYNNMMATPSYMGTFSPEDVPGLAT 746
Db 1569 VEQMTDALLSGQTLVMDSDMRSDQORLYAYMANAVTYLSGTPSVLSLEYGSIPLSR 1628
QY 747 LVLVGQEMSSVNAIWAPKLO-----LLNGYQOSSESSICFASN-----MSTEPNMGRA 796
Db 1629 IDAIGEDFTTPV-----FDKIRSSFGLIINGYGPTE---ISITSHKRLYLKHEP-RLDKS 1680
QY 797 VG-----AHSWVIDPNDINRLVIGAGVELVIESPGIARDYIVPPPPPEKSPFFDTIPSWY 851
Db 1681 IGHVPANTACYVLNP--AMQRPVPGMGELYIGGIVGARYGNRPETLAEFRVSNPQSA 1738
QY 852 PANTFPDGAKYTGDLARYASDGSIVCLGRIDSQVKIRQORVELGAIEHLRQOMPDDL 911
Db 1739 GEKALGNARLYKTGDLVRWLPNGELEYLGRNDMQVKIRQORVELGEVEAILASYPGVTR 1798
QY 912 TIWV----BATKRSQSANSTSLIAFLIGSSYFGNRPSDAHILHDHDKATKINIKLEQVLR 967
Db 1799 ALVLAREYAAASAGGASOKYLVAFYL-----SABELPDTILHWMARLPQAI-- 1847
QY 968 HSIPSYICMELPRTATKIDRRRLRIMGKDILDKOTQGAIVQOAPAPIPVADTAACL 1027
Db 1848 --VPVRIIRILEIPTVSGCKLDVKRL-----PATEFALAGSDYVQPA-----NOTERDL 1895
QY 1028 HSIWQSLGIDPATVNVGATFELGNGNSITATKMV-NMARSVMGMDLKVSNYIOPHTLAGI 1086
Db 1896 CRLWSKVTGLGDRIGSRDDFFGSGGDSLRALKQAQSIHAFQENFNVAFAHPNIAAQ 1955
QY 1087 SAVVK-----GDPLSYTLIPKSTHEGPEQSYSGRLWFLDOLDVGSILWLIPIYAVRM 1139
Db 1956 AAFLENASAGAALSPAGLAAAGAAP-PVSLAQERLLFIDDFVEGTAAYNIPFAPAL 2014
QY 1140 ---RGPVNVNVALRRALAEQRHETLRTTFED-QDGVGVQIVHREKLSSEMKVIDLQGSOL 1195
Db 2015 DSDRAEQ--DAVAAALGVLLQRRHAALRTLQGLHGVQVQNV-LPADEALARFEVRSV 2072
QY 1196 DFEVLNQEQITP-----FNLSEAGWRATLLRLGEDDH--ILTIYVHHIISDGWSDVLR 1249

Db 2073 PDRSALDAELVADFAHVRFLDQELPLRLARLFAVSGDRRSIVVSLVHFHSCDFGWSWIRFR 2132
QY 1250 RDLNOLYSAALKSDKPLSALTPLFIQYSDFAKQKQDFIQ--EKQLNMYWKQKQLDSSP 1307
Db 2133 EELQALLLEGATAQI-----LPLLPATYADFSVWQROQLSDORLDAFDYQWORSLAGWOP 2186
QY 1308 AKITPDARPAALLSGDAGCVHVTIDGELYQSLRFAFCNEHNTTSFVVLAAAFRAAHYRUTA 1367
Db 2187 LQPLDHPARPAQFDYLGREIVDFVDATTCQDLRLVAQTTRTSFSSVLLAAAYLTLLKAYSG 2246
QY 1368 VEDAVICTPTANRRPELEDILIGCFVANTOCMRINDHHDHDTGTLINQVKAITTAFAENED 1427
Db 2247 QSDLIIVTPTANRRHGFDFGVIGFEFVFNLLARTTIDGRGRLLDYLREIGGLVLAQALHED 2306
QY 1428 IPIERVVSALQPGSRDLSSTPLAQLIFAV-----HSQKDLGRFRKFOGLES 1472
Db 2307 LPPEQLVKRLGV-EKDSRRHPVQAVFSLLNKESVDTCWPMRPHAPDDEGR----- 2357
QY 1473 VPVPSKAYTRFDEMFHLFOETDSLKGSVNFADFLFKMETVENNVVYVFEILRNLQGSRT 1532
Db 2358 -----TTAKFDLSATVSETAQGLSVNFTYAASLFDASSVDNIVATYRRLIQEFGR---- 2407
QY 1533 PVSTPLTDCGIVTLEKLDVLNVKHVDYPPRESSLAD-----VFQYQVSAVPSDLAVV 1583
Db 2408 ---LAPWESAAL--SDIRYVEACAAPAAIAEERAAETLHAAFEKIAAANGDDLLAVA 2461
QY 1584 DSSCRITYTDLDRQSDILA-GWLRRRSMPAETLVAVFAPRSCETIVAFFGVKLANLAYLP 1642
Db 2462 FGOTRLSYRELDARANRLARSLLASRTLAPGDFVALVDKSEWTVVAILAVWKAGAAVVP 2521
QY 1643 LDVRSPEARVODILSLSGSTIV-----LIGHTDAPPDIEVTNVEFVRIROALN 1691
Db 2522 IDGYPDERIAFLEDTGARAVIADEASYPRISALAGHAARP-----VLNVH-----A 2569
QY 1692 DSNADGFEVIEHDSHTKPSATSLAXLYTSGTGRPKGMIEHRIIIRTVTSGCIPNYPSE 1751
Db 2570 HASASASELLPAPASAAGDDFAYIYTSOTGRPKAVLVHRNVL-SFAGLQRYFGD 2628
QY 1752 TRMAHMATI-----AFDGASYEYISALLFQRTLCVVDYMTTLDARALKDVFV-REHVNAA 1805
Db 2629 AGAARQAAILFLSSVYVDFSVQELALSILSGNALLIPESLLFD-----DAFYORMNQERL 2683
QY 1806 SHVTSSQDQVPLRVPRLSRTLMEFFFLAVVTDSTAPDALD-AQGLYQGVQCYNGVGPTEG 1864
Db 2684 SYISGTPQQLFDLARKDH--LHAVLVAGEAFRAHFDKIRAEYRG-PLYNAYGTTE-- 2738
QY 1865 VMSTIYPIDSTESFINGVP-----IGRALNNSGAYVVDPEQQLVGIGVMGELVVTGDLAR 1920
Db 2739 --TTVYNL--VKRFAPDAPYRNDLGHAIANTHELHILDGELRELPPGGLGEIYVAGDCVGA 2794
QY 1921 GYSKD-ALDENRV-----HITVNDQTVKAYRTGDRVRYRIGDGLIEFFGRMDTQ 1969
Db 2795 GYLNRLDELSRERFVANPFRSEDDRLGRND---TLYKTGDLVRRR-HSGELEFFGRNDHQ 2850
QY 1970 FKIRGNRIESAETEAALLRDSVSDAAVVLQOENEDQA--PEILGFVVADHDHSEN 2022
Db 2851 VKINGVRIELGEIEAFAAAPPVGAQCAVVARHGDGERSGSRVLCYCYVAAHNVSSES 2905

Search completed: May 30, 2003, 12:44:23
Job time : 162 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:41:41 ; Search time 37 Seconds
(without alignments)
2488.225 Million cell updates/sec

Title: US-09-482-788-2
Perfect score: 16128
Sequence: 1 MEYLTAVDGRDLPPTPASF.....RVEHLLBEVSKTFEGLNSSL 3129

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-----------------------|-------------------|
| 1 | 5000 | 31.0 | 15281 | 2 US-08-471-119A-2 | Sequence 2, Appli |
| 2 | 1740.5 | 10.8 | 3079 | 4 US-09-413-814-80 | Sequence 80, Appl |
| 3 | 1738 | 10.8 | 3072 | 4 US-09-413-814-93 | Sequence 93, Appl |
| 4 | 1516 | 9.4 | 3712 | 2 US-08-222-617A-4 | Sequence 4, Appli |
| 5 | 1515 | 9.4 | 3665 | 2 US-08-222-617A-13 | Sequence 13, Appl |
| 6 | 1515 | 9.4 | 3712 | 2 US-08-222-617A-25 | Sequence 25, Appl |
| 7 | 1425.5 | 8.8 | 2404 | 4 US-09-134-001C-3464 | Sequence 3464, Ap |
| 8 | 1392 | 8.6 | 2628 | 4 US-09-413-814-11 | Sequence 11, Appl |
| 9 | 1342.5 | 8.3 | 3727 | 2 US-08-222-617A-27 | Sequence 27, Appl |
| 10 | 1342.5 | 8.3 | 3727 | 2 US-08-222-617A-2 | Sequence 2, Appli |
| 11 | 1341.5 | 8.3 | 3666 | 2 US-08-222-617A-12 | Sequence 12, Appl |
| 12 | 855 | 5.3 | 739 | 3 US-08-510-646B-33 | Sequence 33, Appl |
| 13 | 782.5 | 4.9 | 1410 | 4 US-09-335-409-3 | Sequence 3, Appli |
| 14 | 782.5 | 4.9 | 1410 | 4 US-09-568-102-3 | Sequence 3, Appli |
| 15 | 782.5 | 4.9 | 1410 | 4 US-09-567-969-3 | Sequence 3, Appli |
| 16 | 782.5 | 4.9 | 1410 | 4 US-09-568-480-3 | Sequence 3, Appli |
| 17 | 782.5 | 4.9 | 1410 | 4 US-09-568-486-3 | Sequence 3, Appli |
| 18 | 782.5 | 4.9 | 1410 | 4 US-09-568-472-3 | Sequence 3, Appli |
| 19 | 782.5 | 4.9 | 1410 | 4 US-09-567-899-3 | Sequence 3, Appli |
| 20 | 715.5 | 4.4 | 797 | 2 US-08-222-617A-9 | Sequence 9, Appli |
| 21 | 662 | 4.1 | 543 | 4 US-09-413-814-91 | Sequence 91, Appl |
| 22 | 662 | 4.1 | 544 | 4 US-09-413-814-81 | Sequence 81, Appl |
| 23 | 657.5 | 4.1 | 768 | 2 US-08-222-617A-5 | Sequence 5, Appli |
| 24 | 657.5 | 4.0 | 798 | 2 US-08-222-617A-8 | Sequence 8, Appli |
| 25 | 634.5 | 3.9 | 758 | 2 US-08-222-617A-6 | Sequence 6, Appli |
| 26 | 588.5 | 3.6 | 2618 | 4 US-09-413-814-28 | Sequence 28, Appl |
| 27 | 526.5 | 3.3 | 822 | 2 US-08-222-617A-7 | Sequence 7, Appli |

Sequence 82, Appl
Sequence 94, Appl
Sequence 42, Appl
Sequence 89, Appl
Sequence 76, Appl
Sequence 44, Appl
Sequence 50, Appl
Sequence 8, Appl
Sequence 22, Appl
Sequence 88, Appl
Sequence 4902, Ap
Sequence 86, Appl
Sequence 82, Appl
Sequence 24, Appl
Sequence 94, Appl
Sequence 18, Appl
Sequence 84, Appl
Sequence 92, Appl

ALIGNMENTS

RESULT 1
US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15281 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-2

Query Match 31.0%; Score 5000; DB 2; Length 15281;
Best Local Similarity 32.4%; Pred. No. 0;

QY 2532 GGHSLLATKLSIDQRUKVRITVKVDHPDHPVADLASVIRG----- 2574
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 13651 GGHSLLATKLSARLSRRLDAGITVKVQDPQVADLAASILQSSRRHSIPSLPYEGPVE 13710
QY 2575 ----- 2574
Db 13711 QSFAGRLWFDQDNIDALWYLPFALRMGRPLQVDALAAALVALEERHESLRTTFEERD 13770
QY 2575 ----- 2574
Db 13771 GVGIVVQPLRTTKDIRIIVSGMRDDAYLEPLQKEQOTPDILASEPGWRVALLKGD 13830
QY 2575 ----- 2574
Db 13831 DHILSIVMHIIISDGWSTEVLQRELQGFYLAASKGAPLSQVAPLIQYRDFAVWQOE 13890
QY 2575 ----- 2574
Db 13891 QVARSOROLDYWKQOLADSPAELLADYTRNVLSGEAGSVFVINDSVYKSLVSFCRSR 13950
QY 2575 ----- 2574
Db 13951 QVTTFTLLAARAHYBMTGDDATIGTPIANRNPELENLIGCFVNTQCMRITIGDDE 14010
QY 2575 ----- 2574
Db 14011 TFSILVQOVSTTATAFENQDVPFIRIVSTLSAGSRDTSRNPLVQLLFAVHSQQGLGRIG 14070
QY 2575 -----LG 2576
Db 14071 LDGVDEPVLSTVTRFDLEPHAFQADRLNGSVMFATDLFPETIOGFVAVVEVLQ 14130
QY 2577 LQQPV----- 2582
Db 14131 LEQOSPATPLAEGIAQLRDAGALQMPKSDYPNASLDVDFQQQAMASPTVAVTDS 14190
QY 2583 ----- 2582
Db 14191 SKLYAELDRSDQAASYLRQQLPAETWAVLAPRSCETIIAFLAILKANLAYMPLDVN 14250
QY 2583 ----- 2582
Db 14251 TPSARMEAIISVPGRRILVSGVRHADINVPNAKTMLSIDVTGTDAITGPELVVRP 14310
QY 2583 ----- 2582
Db 14311 SATSLAYVIFTSGTGKPGVMEHRAIMRLVKDSNVVTHMPATMAHVTHNIAFDVSLF 14370
QY 2583 ----- 2582
Db 14371 EMCATLLNGGTLVCIDYLTLLDSTMLRETFFEREQVRAAIFPPALLRQCLVNMPOAIGMLE 14430
QY 2583 ----- 2582
Db 14431 AVYVAGDRFHSRDARATQALAGPRVYNAYGPTENAILSTIYNIDKHPYVNGVPIGSAYS 14490
QY 2583 -----DGQC----- 2587
Db 14491 NSGAVMDRNOQLPPGVNGELVVTGEGVARGYTDASLDTDFVTVTIDGQRAYRTGD 14550
QY 2588 ----- 2587
Db 14551 RVRVRPKGQIEFFGRLDQAKIRGHVRVELGEVHALLSENSVTDAAVLRTMEEDPQL 14610
QY 2588 ----- 2587
Db 14611 VAVTTDHEYKSGSNEEDPYATQAAGDMKRLSLPYVWPSRVITLROMPLNANGK 14670
QY 2588 -DRS-----AHMAPRTEAIIICDEFKVLGFOVGTIDNFFDLGGHSL 2629
Db 14671 VDRKDLARRAQWTPASSGPHVAPRNETAAICDEFETILGVKVGITDFFELGGHSL 14730
QY 2630 MATKLAVRIGHRLDVTVSVDHFDHPVLFLQALIALDLNV-----QSKTNEIVGGRMAE 2683

Db 14731 LATLAARLSRRMGLRISVKDLDFDPVPVSLAGKLEQQQSGSEDSSTVGIV----- 14783
QY 2684 YSPOLLTTEDEEFEMASEIKPOLEL-QEIIODIYVSTOMOKAFDLHTTARPRFPVEFY 2742
Db 14784 --POLLPAEMSRREIIQDQVVPVPIENGHSTPLDMYPATOTIFFLHKDKATGHPATPPLFS 14841
QY 2743 IDFFSTSPDAAGLIKACESLVNHLIDIFRTVFAEASGELYQVVLSCLDLPQIVIEEDNI 2802
Db 14842 LDFPETA--DCRRLASACAALVQHPDIFRTVFSRGRFYQVVLHLDVPEVIEETEQL 14899
QY 2803 NTATNEFLDEFAKEPVRGLHPLIRTIKQT-KSMRVIMRISHALYDGLSLHHVVRKLMH 2861
Db 14900 DEVALALHEADKQOPLRLGRAMRLTAILKRFCAKMRVLKMSHSLYDGLSLHHVNALHA 14959
QY 2862 LYNGRSLPPHQFSRYMOYTDGREGHGFWRDVQNTPTMTLSDDT-VVDGNDATKAL 2920
Db 14960 LYSOKHLAQAFGLYMHMMASRRAEYNFNRSLQGSMTSLKRSVGALEAMT8SAGTW 15019
QY 2921 HLSKIVNIPSOVLGRSSNIITQATVFNAAACALVLSRESKSDVVFGRVSGRQLPVEYQ 2980
Db 15020 QTSKIRIPPAALK--NGITQATLFTAASVLLAKHTKSTDVVFGRVSGRQDLSINQ 15076
QY 2981 DIVGPTNAVYRAHI-BSSDYNQLLHDIDQOYLLSLPHETIGFSDLRKNCCTDWEATN 3039
Db 15077 DIVGPCINEVPVVRIDEGDGMGLLRAIQDYTSFRHETILGLOEVKENCNTDWTATKE 15136
QY 3040 FSCCITYINFVHPESQFEQORVEM----- 3064
Db 15137 FSCCIAFONLNLHPAEIEGQOIRLEGLPAKQARQANGHPNGTNGTNGTNGANGT 15196
QY 3065 -----GVLTKFVNIEM-----DEPLYDLATAGEVEPDGAGLKVTVIK 3102
Db 15197 NGTNGTNGHANGINGSVNGRDSNVVSAACQAPVHDLDIVGPEPDGS-VKIGIGAS 15255
QY 3103 TOLGGRKVRHLEEVSKTFEGLN 3126
Db 15256 RQILGEKVGVMGMLNLCETMLALS-15279

RESULT 2

US-09-413-814-80
; Sequence 80, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 3079
; TYPE: PRF
; ORGANISM: Sorangium cellulosum
US-09-413-814-80
Query Match 10.8%; Score 1740.5; DB 4; Length 3079;
Best Local Similarity 27.9%; Pred. No. 3.8e-143;

Matches 614; Conservative 348; Mismatches 858; Indels 381; Gaps 83;

QY 15 PTPASFGHSDGSPLNSS-----YEQLPHVLGSDSRTEAKPCPTPFOLD-----MI 60
Db 898 PTLAQLASHLSSGAAATAALERGLTRPDGPPSPRVAT--PEEPFALTGORAMWL 955
QY 61 DCNALDKQSAIGHAVYD-----VPTDIDISREAFALAKKEIVNOTPALRAFAFAFTSDGSKTS 114
Db 956 EC-----QKSADG-ALYNLGRVTRVLGAGVDVAALRAPEGLVERHEARLTFTRDGHPLQ 1010
QY 115 QVILKDSFVSWCMWSSSSP-----DEVVRDEAAAAAGPCRNFRVLEDM 161
Db 1011 QV--HRRVALEW-----AEEFAMALDEREIVARADE--VRRRAFDFLERGPLL-RVHVWRRG 1061
QY 162 QTKKCOLWTFSHALVD--VTQQRVLSRVFAAYKHEKTHRPETPESSDADTDTSQSV 218
Db 1062 EGQPPLLTVVHHLVVDYWSFALLVRELGLYSALRAGR-PPOLPPPS-----SF 1110
QY 219 SVVSMSCD--NAVSAATHFWOTHLN-----DL-----NASVFHLSHLMVNPPTTT 263
Db 1111 FAAGVSCPSPREAAGAEYWRKALDGAATTATIDLPORARHDASPRCRAHAITLPKPLTG 1170
QY 264 AEHRITPLSOKALSNSAICRTALSILLSRYTHSDREALFCGAVTEQSLPFDKHYLDGTYQ 323
Db 1171 ALARLA---RERTGTTLSVLLSALTVLLHRASQNDLVVG-VPSAGRNDDDESTRAFGYFV 1226
QY 324 TVAPLRVHCOSN-----LRASDV-MDAISSYDDRLGHLAPFLGRDIRNTGONGSACD 375
Db 1227 QMLPVRLAEGAASFDALVARVDAFLDGLAHGDSALQHL-----LAEPRGAARRGGALFD 1282
QY 376 -----FOTVL-----LVTDGSHVNNNGINGFLQQTITESHFMPCNNRALLLHCQMSSG 423
Db 1283 VAFQFQALPSLPRLAALTTGAEDVRIAQGELELTTLADEQAAAEFDLALFAELD-SG 1341
QY 424 ALLVAYYDHNVIDSLQTLRLQFQGLHKLKQSPDLSS--MAEVLNMTYEDRA-EIESW 480
Db 1342 IALRFYDQOLFDPATIERMAR---HFVLLLESAVEHPCGRPLSELRLMSLDAERALLDDW 1398
QY 481 NS-----QPLEVODTLIHHEMLKA-VSHSPTKTAIOAGDGDWTYSELDNVSSRLAVHKS 534
Db 1399 SGAAAAQAASAPACVHALFEAHAAROPDATALEFGHQRTYAEIWSLSTWELALWLKD 1458
QY 535 LGLRAQOALIPVYFKSKWIVASMLAVLKSNAFTLIDPNPPARTAAQVVTOTRATVALT 594
Db 1459 RGV-GPGSVGVGCIERSPRWAAQLAVLKAAGAYASLDPANPPARLAEMLADCAALVLT 1517
QY 595 SKLHRETQKLVGR-CVVVDDELLO--SVSASDDFSLSLKSDQLAVITFSGSDGPKGI 651
Db 1518 SS---QASHKLTAAAPCV---HLVODGACAPSTHPLVSRPDDLAYVLFVTSGGSGTPKGV 1571
QY 652 MIEHRAFSSCALFKFASLGINSDFRALQFTHAFGACILLEIMTLLINGCGVCIPSDDDRM 711
Db 1572 CVRHASLSRLVSFFQHLALSPPDRWTQLASSGFDASVYEIWTPLACGAALLL-ADDDAL 1630
QY 712 NSIPSPINRYNNMM-----ATPSYMGTFSP-----DVPGLATVLV-VGEQMSSSVNAIW 762
Db 1631 RS-PTAL-----VSLVAQATLSFMTPLAEACFEQDWTGIALRAMTVG--GDKLHPLR 1682
QY 763 AP-KLQLLNGYQSES-----SSICFASNMSTENNNGRAV-GAHSWVIDPDNDINRLVPIGA 817
Db 1683 RPLPRLFNWYGTEATVITVAEADLGDPE-PLGRPIDLSALVYVLDPH--MQPVPGV 1739
QY 818 VGELVIESPGIARDYIVPPPEKSPFFTDIPSWYANTFPD-----GAKLYRTGDLARYA 872
Db 1740 LGELYIGGACLAQY-----RDLTA---ERFLPDFGQGARLYRTGDLVRWR 1786
QY 873 SGGSIVCLGRIDSQVKIRQORVELGAIETHLRQOQPDLLTIVVEATKRSQSANSTS-LIA 931
Db 1787 PDGOLAFAGRDRQVKLGRVRELGEVESALRR-----LPVREGVVLHGGOSAARLIA 1841
QY 932 FLIGSSYFCGNRPSDAIHDHDKATKAINIKLEOVLRPHRSTPSYICMLPELPRATATCKIDRR 991
Db 1842 YVPGCA-----DPPSERDLRGMARLVDPALVPAHFVLLPALPMSLSGKVDK 1889

QY 992 RLRLMGKDILDKQFOGAIVOQAAPAPIPVFAD-----TAAKLHSIWQSLGIDPATVNV 1044
Db 1890 LL-----PAPPAAHADYEPPEGELELRELAHIQSVLHLD-----RV 1925
QY 1045 GA--TFPELGGNSITAIKMW-NMARSVGMMDLKVSNIYQHTPLAGISAVVKGDDPLSYTL-- 1099
Db 1926 GRHDSFFDLGGHSLAMQVLGRSIESSUGIRTTLTFLFEHTLTHQADRLSSGAASSTAAA 1985
QY 1100 -----IPKSTHIEGPPVEQ-----SYSQGRLWFLDQLODVGSLWYLIPIYAVMRGPNVNDAL 1148
Db 1986 ATPASEIAPSLGRAPADEPYPLSYEQERLWVLEQLLPGGTAYNVVQAVRLNRNLVDVAL 2045
QY 1149 RRALAALQORHETLRTTFEDODGVGVQIVHEKUSEEMKVIDLCCSLDLPFEV-----LNQ 1203
Db 2046 SSALAALVRHRSLSLRTVFV-ASPTPAQKICEPEAAPAEVVDLRTGTPDEAAAAARAWASR 2104
QY 1204 EQTTPFNLSSEAGWRATLLRLGEDDHILTIVMHIIISDGHWSIDVLRLQNLQYSAALKDS 1263
Db 2105 EQATGFDLARGPVFERARLFRDLHDHVCVLVLTSHIIVTDAMSFPQPLVRLDLAELYRRARGG 2164
QY 1264 KDPLSALTPLPQIYSDFAKWKQDFIQE--KQINWKKQKDKSSPAKIPTDFARPALLS 1321
Db 2165 P---ADMPELPLQYVDFAVWQRRHLAGKRLADKLAWHTATLRLGLPVLELQTDPRPPVQT 2221
QY 1322 GDAGCVHTTIDGELYQSLRAFCHNEHTTSFVLLAAPRAAHYRLTAVEDAVIGTPIANRN 1381
Db 2222 FGAERVLPLDARLVAQLDELARSRGATRFPMVLLAALGVLLRRSSGQDDDLAIGTAVANRP 2281
QY 1382 RPELDTIGCFVNTQCMRINIDHDTFGTLINOVKATTTAAFNEDIPFERVVSALOPGS 1441
Db 2282 RPELEPLGVFFVNTIVMRDLGDDPTFEELLSPARKVALEAFEHQDAPFEKVVEAVNP-R 2340
QY 1442 RDLSSSTPLAQLIPAVHSQK----DLGRFKQGLSVPPVPSKAYTRFDMFEHLFQETDSLK 1497
Db 2341 ROLRSPLQVLMVQNAPEALELGEVRIEPLD-LPVEA--TRDLRSVEPRGGRDV 2396
QY 1498 GSNFEADELFKMETVENVVRVFEILRNGLQSSRTPVSIPLTLD-----GIVLEKLDVL 1552
Db 2397 ISQYVNDLFDAAITDIRMLATMOSVLSRATQDPAQVRVRLSVAPEDRERALVAMNDTAVA 2456
QY 1553 NVKHVDYPRESSLADYFQVSAYPDSLAVDSSRLTYTDELQSDILAGWLRRRSMPA 1612
Db 2457 TPDHL-----RLEEFFEFERAVEQDPAQVAVDAERLITGELARRAEIAAASRSRGATA 2510
QY 1613 ETLVAVFAPRSCETIVAFGLVKANLAYLPDVRSPSARVQDILSLGSLGPTIIVLIGHDTA 1672
Db 2511 NALVAVVMEKGWEQVAALVGLVRAGAAYLPLDPLRPEERLHLL----- 2554
QY 1673 PPDIEVTNVFEVRIRDALNSNAD-----GFEVIEHDSK-----PSATS 1712
Db 2555 -----EHAERVLVLTQSAYVDGTIAMPAGIERLAVDADERWREQPVARRPPGGSTDD 2605
QY 1713 LAYVLTSGTGRPKGMIEHVRVIRVTS-----GCIPNYPSETRAHMAHATIAFDGA 1765
Db 2606 LAYVIYTSSTGLPKGMIDHRCNAVTVLIDINRRFDVG-----PEDRVLAUSLSFDLS 2659
QY 1766 SYEISYALLGFRVLVCVDYMTTLDARALKDVFFREHVNAAASHVTS----- 1810
Db 2660 YVDVEGTLAAGGAVVIPOTRASDPCHWRELVERERVTVNWSVPALMEMLMDASPCGADP 2719
QY 1811 -----SSQDVPRLVRPRLSRTLMEFFLVVTDSTAPDALDAOGLQGVOCYNGYGP 1860
Db 2720 ALUSSRLVMSSGDIPLKLPDIRAA-----CRAPRVVSLG-----GA 2757
QY 1861 TENGVMSTIYPIDSTESFINGVIGIRALNNSGAYVVDPEQQLVIGVMGELVVTGDLAR 1920
Db 2758 TEASIWSTIAHPADVDPAWRSPYGPPLANQHTYVLDGLEPCPGVGPVEIHIGGICVAL 2817
QY 1921 GY-SKALDENRFV-HITVNDOTVKAIRYTGDRVRVIRIGDGLIEFFGRMDTQFKIRNRIE 1978
Db 2818 GYWRDEARTRERFLKHPTTGE---RLYRTGDLGRV-FADGTIELLGRTHQVKIRGRIE 2873

Db 2334 RLSRSLPQVLMVQVNAPEALELGEVRIEPLD-LPVEA---TRFDLRFSEVRGRGDV 2389
Qy 1498 GSVNFAELFKMETENVVVRFFBILRNGQSSRTPVSLPLTD-----GIVTLEKLDVL 1552
Db 2390 ISLOYNVDLFAAATIDRMLATMSQVLSRATQPAQRVRLSVAPEDRERLAVNMDTAVA 2449
Qy 1553 NVKHVDVPRESSLADVQTVSAVPDSLAVVDSRCLTYTFLDQSDILAGLWLRSSMPA 1612
Db 2450 TPDHL-----RLEPFFERAVEQDACAVVDAERRUTYELARRAIAAASRGATA 2503
Qy 1613 ETLVAFAPRSCETIVAFGGKLANLAYPLDVRSPSARVQDILSGSLGPTIVLIGHDTA 1672
Db 2504 NALVAVVMEKWEQVAAVLGVRAGAAYPLDPLPERLRHL-----2547
Qy 1673 PPDIEVTVFVRIRDALNSAD-----GFEVIEHDSK-----PSATS 1712
Db 2548 -----EHAERVLVTQSAVDGTTIAPAGIERLAVDADDERWREQPVARRPPGGSTDD 2598
Qy 1713 LAVLYTSGTGRGPKWIEHRVIRTVTS-----GCIPNYPSETRMAHMATIAFDGA 1765
Db 2599 LAVVYTSSTGLPKGMIDHRGAVNTVLDINRFDVG-----PEDRVIALSLSLFDLS 2652
Qy 1766 SYEYISALLFGRTVLVCVDMYMTTDLARALKDVFREHVNAASHVTS-----1810
Db 2653 VYDFEGTLAAGAVVIDRTPASDPGHRELVERERVTVNSVPALMEMLMDASPGAGDP 2712
Qy 1811 -----SSQDPLRVPRRLSRLTLMFFLVVTDSTADALDAQGLYQGVQCYNGYP 1860
Db 2713 ALLSLRLVMSGDWIPKLPRIRAA-----CRAPRVSLG-----GA 2750
Qy 1861 TENCVMSTIPIIDSTESFINGCPVIGRALNNSGAVVDPQOOLGIGVYMGELVVTGDGLAR 1920
Db 2751 TEASIWTAHPADVDPANWSIPYGRPLANOHTVLDDEGLPCPIGVGEIHGGIGVAL 2810
Qy 1921 GY-SDKALDENRFV-HITVNDQTVKAYRTGDRVRYIGDGLIEFFGRMDTQFKIRGNRIE 1978
Db 2811 GYNREARTREFLKHPITGE--RLYTGDLGRY-FADGTIELLGRTHQVKIRGFRIE 2866
Qy 1979 SAEIEAALLRDSVRDAVVLQONEDQAEILGEVW-ADHD 2018
Db 2867 LGETEAAALQHPSEVQAVAAKTDPSGKRLVAVVVGADGD 2907

RESULT 4

US-08-222-617A-4
; Sequence 4, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222.617A
; FILING DATE: 04-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97,157
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3712 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 2555
; OTHER INFORMATION:
; OTHER INFORMATION: /note= "Xaa-Ala or Ser"
; US-08-222-617A-4

Query Match 9.4%; Score 1516; DB 2; Length 3712;
Best Local Similarity 26.1%; Pred. No. 4e-123;
Matches 564; Conservative 377; Mismatches 912; Indels 310; Gaps 80;
Qy 2 EYLTAVDGRQDLPTTPASFCSH-----GDSPLNSSYEQLPHLYGLDSSRIEAIKPCPTPF 55
Db 866 KYTKASNGTNGVTNGTAHVNGHAANGHVSDSYVASSLQGGFVYHSLKNEUSE-----917
Qy 56 QLDMDICNALDKQGAIGHAVYDVPTDIDISRFALAWKEIVNOTPALRAFAFTSDSGKTSQ 115
Db 918 -----AYTMQSMI-H--YGVPLKRDY--YQAQWQVQGEHFAIR-LRTWEAEVQMI 963
Qy 116 VILKDSVFESVMCWNSSSPD-----EVRDEAAAA-----SGPRCHRFVLLDMOTKK 165
Db 964 VDPKSELDRVVDVTDVSSREKQLVALEQLQTEDLAKVYHLDKGPLMLRYLIL--LPDSK 1021
Qy 166 COLVMTESHALVDVTFQORVLSRVFAAYKH-----EKD-----THRPETP 205
Db 1022 YSLFSCCHHALDQWNSLPLFNNVHQAYLDLVEGTASPEQDAYLYLGGQYQLQSHR-----1077
Qy 206 ESSDATDTSQSVSVSMSCDNVAVSATHFWQTHLNDLNAVFPPLSHLSDHLMVNPPTTAE 265
Db 1078 --DDHLDFAEQIGRIEERCDMNA-----LNEASRYKVP-LADYDQV--REQRQ 1122
Qy 266 HRTTFP-----LSOKALNSAICRTALSIILLSRYTHSDEALFG-AVTEQSLPF 312
Db 1123 QTISLPWNNSMDAGVREELSSRGITLHSLQTVWHLVHLSYGGGTHITGTITISGRHLVP 1182
Qy 313 DKHYLADGTYTQTVAPL--RVHCOSNLRASVDMAISSYDDDRGLHGLAPFGLDRINTGDN 369
Db 1183 PGTERSGLFINTLPMIFDHTVCOD---MTALGAIEHVQGVQVNMNMRGHWELGRMSKN 1238
Qy 370 GSAACDFQTVLLVTDGSHVNN-----GINGFLOQITESSHFMPCCNNRALLHC 417
Db 1239 DLKHGLFDTLVLENYPNLDTQEREKHEEKLFKTKGTEKLSYP-----LAVIAQE 1290
Qy 418 QMESSGALLVAY-----YDNVIDSLQTLRLQOFGHLIKCLQSPDLSSMAEVLNMTPEY 472
Db 1291 DGDGSCFTLCYAGELETFDESIALDLTVR--DTLSIDLGINHAPI-----RNMELSSN 1343
Qy 473 DRABIESWNSQPLEVQDTLHHEMLKAVSHSPKTAQANDGMDTYSFLDNVSSRLAVHI 532
Db 1344 QTAQDKWNATAFEPYPTTLHAMFESAQQPKDVAVVYEDIRLTYRELNSRANALAPYL 1403
Qy 533 KSLGLRAQQAIIPIVYFEKSKWVIASMLAVLSKSGNAFTLIDPNDDPARTAAQVVTOTRATVA 592
Db 1404 LSQAIIQPNKLVGLIMDKSEHMTSILAVNKTGGAYVPIDPRYDQRIQYILEDTAALAV 1463
Qy 593 LTSKLHRETQKLVGR--CVVVDDELQSVASDDFSLTKSQDLAVVIFTSSTGSDPKG 650
Db 1464 ITDSPHIDRLRSITNNRLPVIQSDFALQLPSPVHPVSNCKSPKSLAYIMYTSSTGNPKG 1523
Qy 651 IMIEHR-----AFSSCALKFGAGINSDDTRALQFGTHAFGACALLEMTTLINGGCVCIP 705

; FEATURE:

; NAME/KEY: Protein
; LOCATION: 1..3665
; OTHER INFORMATION: /label= ACYS
; OTHER INFORMATION: /note= "ACV Synthetase from Acromonium
; OTHER INFORMATION: chrysoeum; aa 1-3665"
US-08-222-617A-13

Query Match 9.4%; Score 1515; DB 2; Length 3665;
Best Local Similarity 26.1%; Pred. No. 4.8e-123;
Matches 564; Conservative 377; Mismatches 912; Indels 310; Gaps 80;

QY 2 EYLFAVGRQDLPTTPASFCSH-----GDSPLNSSYEQLFHLGCLSSRIEAIKPCPPF 55
DB 866 KYTKASNGTNGVTNGTAHVNGHAANGVSDSYVASSLQOGFVYHSLKNLSE----- 917
QY 56 QLDMDICNALDKQSAIGHAVYDVTDIDISRFALAKWEIVNQTPALRAFAFTSDSGKTSQ 115
DB 918 -----AYTQSMI-H-YGVPLKRDY--YQAAQVRQGEHPALR-LRFTWEAEVQI 963
QY 116 VILKDSFVFSMCMSSSSSPD-----EVVRDEAAAA-----SGPRCNRVLLLEDQTKK 165
DB 964 VDKSELDRVVDVTDVSSREKQVLALEQLOTELDKAVYHLDKPLMLRLYLIL--LPDSK 1021
QY 166 COLVMTSHALVDVTFQORVLSRVFAAYKH-----EKD-----THRPETP 205
DB 1022 YSCLFSCHHAILDGSWPLFLFNVHQAYLDLVEGTASPVQDATYLLGQQYQLOSHR---- 1077
QY 206 ESSDATDTSQSVSVVSKCEDNAVSATHFWOHLNLDNASVPHLSHDHLMVNPNTTAE 265
DB 1078 --DDHDFWAEQIGRIEERCDMNL-----LNEASRYKVP-LADYQV---RERQ 1122
QY 266 HRTTFP-----LSQKALSNAICRTALSILLSRYTHSDEALFG-AVTEQSLPF 312
DB 1123 QTISLPWNMSDAGVREELSSRGITLHSILQTVHVLVHSGYGGTHTITGTTSGRHLVP 1182
QY 313 DKHVLADGTVTVAPL---RVHCQSNLRASDVNDIAISSYDDRLGHLAPFGLRDIRNFGDN 369
DB 1183 PGTERSGLFINTLPIDFHTVQD---MTALEATEHVQGVQVNMNRSGVNLGRMSKN 1238
QY 370 GSAACDQTVLLVTDGSHVNN-----GINGFLOQITVESHFPCNRRALLHC 417
DB 1239 DLKHGLFDLFLVLENTPNLDEQREKHEKLTIRKGTGKTSYP-----LAVIAQE 1290
QY 418 QMESSGALLVAY-----YDHNVDLSLQTTLLQOFGHLIKCLOSLDLSMAEVLNMTY 472
DB 1291 DGDSGCSFTLCYAGELFTDESIGALLDTR--DTLSIDLGNHAPI-----RNKEYLSSN 1343
QY 473 DRAIESWNSQPLEVQDTLIIHMLKAVSHSPKTAIQAWGDWMTYSELDNVSRLAVHI 532
DB 1344 QTAQLDKWATAFEYPTNTLHAMFESFAQQPKDKVAVVYEDIRLTVRELNSRANALAFYL 1403
QY 533 KSLGLRAQQAIIIPVYFEKSKWVIASMLAVLKSNAFTLLDPNDPPARTAQVVTQTRATVA 592
DB 1404 LSQAATOPKNKVLGIMDKSEHMTSLIAVWKGTGGAIVPIDRYPDQRIQYILEDTAALAV 1463
QY 593 LFSKLHRETQVLVGR--CVVVDDELQSVASDDFSSLTQSODLAYVIFTSGTGDPKG 650
DB 1464 ITDSDHIDRLSRITNRLNPVIOQDFALPPSPVHPVSNCKPDLAYINVTSGTGNPKG 1523
QY 651 IMEHR-----AFSSCALAFGLSNGSDTRALQFQTHAFGACLLIMTTLINGGCVCIP 705
DB 1524 VWEHGGVNVLCVSLCL--FG--LRNTDDEVILSFNSYVDFHFEVQMTDALLNGQTLVVL 1580
QY 706 SDDDR--MNSIPSFIRNVNMMATPPSYMGTFSPEDV--PGLATVLVGEQSSSSYNATW 762
DB 1581 NDEMGRDKERLYRIETNRVTVLSGTPSVISMVEFDRFDHRLRRVDCVGEAFSEPY---- 1636
QY 763 APKLO-----LLNGYGOSESSICFASNMSTEPNNMRAVGAH-----SWVIDPNDINR 811
DB 1637 FDKIRETFGLIINGYGPTEVSIITHKRPYPPERRTDKSICQQLDNSTSYVL--NDDMK 1694
QY 812 LVPICAGVELVIESPGIARDYIVPPPEKSPFFTDIPSWYPANTFPD-----GAKLY 863

DB 1695 RVPICAVGELYLGGDVARGY-----HNRDPLT--ADRFANPFQTEOERLEGRNARLY 1746
QY 864 RTGDLARYA-----SDGSIVCLGRIDSQVKIRGORVELGAIETHLQQQPPDDLTIVVEATK 919
DB 1747 KTGDLVRHNGANGDETEYIGRNDQFVKIRGORIELGEIEAVL--SSYPGIKSQSVVLAKD 1805
QY 920 RSQSANSTSLIAFLIGSYFNGRPSDAHILDHDAIKAINIKLEQVLPRHSIPSYFICMLE 979
DB 1806 RKNDGQK-----YLVG---YF---VSSAGSL---SAQAIRREMLTSLPDYMYPAQLVPTAK 1852
QY 980 LPRATGKIDRRRLRIMGKDILDKQTQCAIVQQAIPAPIPVADTAAKLHISWQSLGIDP 1039
DB 1853 FPVTVSGKLDKALPVPDPTVEDD-----IVPRTVERILAGIWESELLEIPV 1900
QY 1040 ATVNVGATFFELGNSITAIKWVNA-RSVGMDLKVSNIYOHPTLAGIS-----1087
DB 1901 DRISIDYFFSLGDSGLSKLSTKLSAATPALGAVAVSRNLFSPHTIEALUSQMIIRGSNEVK 1960
QY 1088 --AVVKGD-PLSYTLIPKSTHEGPVQSYSGRLWFLDQL-----DVGSLWYLIPIYAVRM 1139
DB 1961 DVAVVKGASLDIPLSP-----AQERLMFIHEFGHSGEDTGA--YNVPLQQL 2006
QY 1140 RGPVNDALRRALAALEORHETLRT--TFEODDGVQVQ-IVHEKLSSEMKVIDL--CGSD 1194
DB 2007 HHVCLLESLEKALRDVSRHEALRTLITRTQSSVHCOKILDAEAOQLFSVDVLRLTSE 2066
QY 1195 LDPEVLNQEQTTTPNLSSEAGWRATLLRLGEDDHIL---TIVMHIIISDGSNIDVLRD 1251
DB 2067 TEMOGRMAESTAHAFKLEELPIHVRLYQVVVROGRTLSFASIVCHHLAFDAWSDVQFD 2126
QY 1252 LNOLYSALKDSKDPLSALTPLPTQYSDFA-KWOKDOFIOEKOL-NYWKOLKDSGPAK 1309
DB 2127 LDFAVHTKHK--AAANLPTLRVQYKEYAIEHRRALRAEQRHVLADYWLKLSMDMEASY 2184
QY 1310 IPTDFARPALLSGDAGCVHVTIDGELYQSLRAFNCNEHTTTSFVLLLAFAHRAHRLTAVE 1369
DB 2185 LVDRPRPAQFDYTGNDLQFSTTPTTAQLKELAKREGSSLYTVVAAAYFLLLVYVYNQR 2244
QY 1370 DAVGTPIANRPELEDIIGCFVNTQCMRINIDHDTFTGLINQVKATTATAFENEDIP 1429
DB 2245 DITIGTPVAHRNHPDESIVVGVFNLLPLRVNVQSODIHG-LIOAVOKELVDAQIHQDLP 2303
QY 1430 FERVVSALQPSRDLSTPLAQLIF-----AVHSOKDLGRFKFQGLSVVPVSKAYTR 1482
DB 2304 FOEITKLLHV-QHDPGRHPLQLQAVFNWNPANVHVEQLQYK---PPSPLPSAA--K 2356
QY 1483 FDMFEHLFOETSLKGSVNFADLFKMETVENVRVFEFLRN-GLOSSRTPVSILPLTD 1541
DB 2357 FDLNVTVKESVNSLVNFNYPSTLSFEETVOGFMETFHLLRLQLAHNKASTLSKLSVED 2416
QY 1542 GIVTLEKLDVLYVKHYDYIPRESSLADVFOQVSAYPDSLAVVDSRCLTYTEL-DRQSDI 1600
DB 2417 GVLNPEP---TNLQPSRDSGNSLHGLFEDIVASTPDRIAADGTRLSVSELNERANQL 2473
QY 1601 LAGLWRRRSPAEVLVAVAPRSCETIVAFGVKLANLAYPLDLDVSPSARVODILSGLS 1660
DB 2474 VHLIISASIVADRLALLDLSIDMVIALLVAKAAGVYVPLDPTYSQRTLEIIESS 2533
QY 1661 GPTVLIGHDTAPPDIETVNVFEVRI--RDALNDSNADGFEVTEHSTKPSATSLAVLY 1718
DB 2534 ARTLIITRKHT-PRGCTVANVSVVLDSLPETLACLNQOKENPTTSTQKPS--DLAYVIF 2590
QY 1719 TSGSTGRPKGVMEHRVITRTVTSGCIPNYPSETRMAH-----MATIADFCASEYISALL 1774
DB 2591 TSGTGTGPKGVLEHQSVVQ-LRNSLIERYFGETNGSHAVFLSNVYVDFSLBQLCLSVL 2649
QY 1775 FORTLVCDVDMTTLDAARALKDVFHREHVNAASHVTSSSQDVPL-RVPRRLRSLTMFFFLV 1833
DB 2650 GGNKLI-IPPEGLTHEAFYDGRREKLSYLSGTSPVLOQIELSLRPLH-----LHMVTA 2703
QY 1834 VTDSTAPDALDAQLYQGVQCYNGYGTENGVMSTIYPIDSTESFINGVPIGRAL---N 1889

| | | | |
|--|------|---|------|
| Db | 2704 | GEEFASOFEKMRSFAG-QINNAVIGITE-----TTVYNIIT--FKGDAPFTKALCHGIP | 2756 |
| QY | 1890 | NSGAYVPEQOLVGIGVGMELVTCGDIARGY-SDKALDENRFV-----HITVNDQT | 1941 |
| Db | 2757 | GSHVVVLDRLQVRPFNAVAGELYLGDCIARGYLNQDALTNERFIPNPFYEPKQASDSRP | 2816 |
| QY | 1942 | VKAYTGRVRYRIGDGLIEFFGRMDTQFKRGNRIESAEIEPAALLRSSVYRDAAVLQO | 2001 |
| Db | 2817 | ORLYKTGLVLRFR-GPHHLEVLGRKDOQVKLGRFRIELSEVRDAVLAIKSAVKAAPIKY | 2875 |
| QY | 2002 | DED 2004 | |
| Db | 2876 | DED 2878 | |
| RESULT 6 | | | |
| US-08-222-617A-25 | | | |
| ; Sequence 25, Application US/08222617A | | | |
| ; Patent No. 582879 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Veenstra, Annemarie E. | | | |
| ; APPLICANT: Martin, Juan F. | | | |
| ; APPLICANT: Garcia, Bruno D. | | | |
| ; APPLICANT: Gutierrez, Santiago | | | |
| ; APPLICANT: Barredo, Jose L. | | | |
| ; APPLICANT: Von Doehren, Hans | | | |
| ; APPLICANT: Palissa, Harriet | | | |
| ; APPLICANT: Van Liempt, Henk | | | |
| ; APPLICANT: Montenegro, Eduardo P. | | | |
| ; TITLE OF INVENTION: A Method for Influencing Beta-Lactam | | | |
| ; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large | | | |
| ; TITLE OF INVENTION: Quantities of ACV Synthetase | | | |
| ; NUMBER OF SEQUENCES: 27 | | | |
| ; CORRESPONDENCE ADDRESS: | | | |
| ; ADDRESS: McDonnell Boehnen Hulbert & Berghoff | | | |
| ; STREET: 300 South Wacker Drive | | | |
| ; CITY: Chicago | | | |
| ; STATE: Illinois | | | |
| ; COUNTRY: USA | | | |
| ; ZIP: 60606 | | | |
| ; COMPUTER READABLE FORM: | | | |
| ; MEDIUM TYPE: Floppy disk | | | |
| ; COMPUTER: IBM PC compatible | | | |
| ; OPERATING SYSTEM: PC-DOS/MS-DOS | | | |
| ; SOFTWARE: Patent In Release #1.0, Version #1.25 | | | |
| ; CURRENT APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US/08/222,617A | | | |
| ; FILING DATE: 04-APR-1994 | | | |
| ; CLASSIFICATION: 435 | | | |
| ; ATTORNEY/AGENT INFORMATION: | | | |
| ; REFERENCE/DOCKET NUMBER: 97,157 | | | |
| ; INFORMATION FOR SEQ ID NO: 25: | | | |
| ; SEQUENCE CHARACTERISTICS: | | | |
| ; LENGTH: 3712 amino acids | | | |
| ; TYPE: amino acid | | | |
| ; TOPOLOGY: linear | | | |
| ; MOLECULE TYPE: protein | | | |
| US-08-222-617A-25 | | | |
| Query Match | | | |
| Best Local Similarity 9.4%; Score 1515; DB 2; Length 3712; | | | |
| Matches 564; Conservative 377; Mismatches 912; Indels 310; Gaps 80; | | | |
| QY | 2 | EVLTAVDGRQDLPPTPASFCSH-----GDSPLNSSYEQLFHYGLDSSRIEAIKPTPF | 55 |
| Db | 866 | KYTRASNGTNGVTNCTAHVNGHAANGHVSDSVASSLQGGFYVYSLKNELSE----- | 917 |
| QY | 56 | QLDMIDCNALQKQSAIGHAVYDPTDIDSRPALAKWEIVNQTALPAFAFTSDSGKTSQ | 115 |
| Db | 918 | -----AYTMQSMI-H--YGVPLKREDI--YQAAWQRVQGEHPALR-LRFTWEAEVMQI | 963 |
| QY | 116 | VILKDSFVFSWMCWSSSSPD-----EVRDEAAAAA-----SGPRCNRFVLLDMDQTK | 165 |

| | | | |
|----|------|---|------|
| Db | 964 | VDPKSELDMRVVWDVTSRSREKQVLEALQLOTEDLAKVYHDKGLPMLRDLYLIL--LPDSK | 1021 |
| QY | 166 | COLVWTFSHALVDVTFQQRVLSRVFAAYKH-----EKO-----THRPETP | 205 |
| Db | 1022 | YSLFSCSHAILDGHSLPLLFNNVHOAYLDLVEGTASPVQDQATYLLGQYQLQSHR---- | 1077 |
| QY | 206 | ESSDATDSDQSVSVMSCEDNAVSAHFQWTHLNDLNASVPHLSHMLMNPNTTTAE | 265 |
| Db | 1078 | --DDHLEWAEQIGRIEERCNMAL-----LNEASRYKVP-LADYDQV---REQRQ | 1122 |
| QY | 266 | HRITFP-----LSQKALSNAICRTALSTLLSRYTHSDEALFG-AVTEQSLEP | 312 |
| Db | 1123 | QTISLPWNNSMDAGVREELSSRGITLHSLQVFWHLVLSYGGGTHITGTTISGRHLPV | 1182 |
| QY | 313 | DKHYLADGTYTQVAPL---RVHCQSNLRASDVMDAISSYDDRLGLHGLAPGLDIRTGDN | 369 |
| Db | 1183 | PGIERSVGLFINTLPMIFDHTVCQD---MTALEALEHVQGVQVNAAMSGNVELGRMSKN | 1238 |
| QY | 370 | GSAACTDFTVLLVTDGSHVNN-----GINGFLOQITESHFPCNNRALLHC | 417 |
| Db | 1239 | DLKHGLFDLFLVLENYPNLDTQREKHEEKLKFTIKGGTEKLSYP-----LAVIAQE | 1290 |
| QY | 418 | QMESSGALLVAY---YDHNVIDSLQTRLLQQFGHLIKCQSPSLDLSMAEVLNLTVEY | 472 |
| Db | 1291 | DGSGCSFTLCYAGELFTDESQALLDTVR--DTLSDIILGNHAPI-----RNMEYLSN | 1343 |
| QY | 473 | DRAETESWNSQPLEVDQDTLIHHMLKAVSHSPKTAIAQWGDWITYSELDNVSSRLAVHI | 532 |
| Db | 1344 | QTAQLDKWNATAFEYPTNTLHAMFESEAQQPKDVAVVYEDIRLTYRELSRANALAFYL | 1403 |
| QY | 533 | KSLGRQAQAIIPIVFEKSKWIASMLAVKSGNAFTLIDPNDPPARTAAQVVTQTRATVA | 592 |
| Db | 1404 | LSQAAIQPNKLVGLIMDKSEHMTTSLAVWKTGGAYVPIDPRYPDQRIQYILEDTAALAV | 1463 |
| QY | 593 | LFSKLHRETVOKLVGR--CVVVYDDELQSVASDDFSSLTQSLQLAYVIFTSGSTGDPKG | 650 |
| Db | 1464 | ITDSPHIDRLRSITNRLNRPVIOQDFALQPLPPSPVHPVSNCKPSDSLAIYMTSGTGNPKG | 1523 |
| QY | 651 | IMIEHR-----AFSSCALKEFGASLGINSDFALQFGTHAFGACLLIEIMTLINGGCVCIP | 705 |
| Db | 1524 | VWVEHGVVNLVCSLCRL-FG--LRNTDDEVILSFSNYVDFHFVEQMTDALLNGTLVVL | 1580 |
| QY | 706 | SDDDR--MNSIPSFNRYNVNMMMATPSYMGTFSPEDV--PGLATLVLVEGEQSSSYNAIW | 762 |
| Db | 1581 | NDEMRGDKERLYRIETNRVTYLSGTPSVISMVYDFDRDHLRDLVDCVGEAFSEPY---- | 1636 |
| QY | 763 | APKLO-----LLNGYGOSESSICFASNMSTEPNMGRAVGAH-----SWIDPNDINR | 811 |
| Db | 1637 | FDKIRETFPCLIIINGPTEVSTITHKRPYPPPERTDKSIGCQLDNSTSYVL--NDDMK | 1694 |
| QY | 812 | LVPIGAAGELVIESPGIARDYIVPPPEKSPFFTDIPSWYPANTFPD-----GAKLY | 863 |
| Db | 1695 | RYPIGAAGELYLGGDGVARGY-----HNRPDLT--ADRFANPFTQERLEGRNARLY | 1746 |
| QY | 864 | RTGDLARYA-----SDGSIVCLGRIDSQVIRGORVELGAIEHLRQOMDDTLIVVEATK | 919 |
| Db | 1747 | KTGDLVRIHNGANGDEIEYLGRLNDFQVIRKIRGRIELGEIEAVL--SSYPIKQSVYLAKD | 1805 |
| QY | 920 | RSQSANSTSLIAFLICSSYFGNRPSDAHLTDHATKAIKILEQVLPRESIFSYICMLE | 979 |
| Db | 1806 | RKNDQOK-----YLVG--YF-----VSSAGSL---SQAQIRRFMLTSLDTPVWPAQLPIAK | 1852 |
| QY | 980 | LPRTATGKIDRRRLRMGKDILDKOTQGAIVQQAAPIPVFPADTAALKHSIWVQSLGIDP | 1039 |
| Db | 1853 | FPVTVSGKLDKALPVPDQTVEDD-----IVPPTVEVERILAGINSELLEIPV | 1900 |
| QY | 1040 | ATVNVGATPFELGGNSITAIKVMNMA-RVGMMDLKVSNIYQHTPLAGIS----- | 1087 |
| Db | 1901 | DRISIVSDPFLSGGSLKSTKLSFAATRALGAVSVRNLFSHPTIEAESQWIRGSNEYK | 1960 |
| QY | 1088 | --AVVKGD--PLSYTLIPKSTHEGVPQSYSGRLNFDLQD-----DVGSLWLVLIPVAVRM | 1139 |
| Db | 1961 | DVAVVKGASLDIPJLSP-----AQERLMFIHEFGHSGEDTGA--YNPVLOQL | 2006 |

```

1140 RGVNVDALRALAALQRHETLRT--TFEDQGVGVQ-IVHEKLSBEMKVIDL--CGSD 1194
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2007 HHVCLSEKALRDVVSRRHEALRTLTRTKSSVHCQKILDAEEAQKLFSDVRLRTSE 2066
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1195 LDPEVLNQEQTTFFNLSSAGWRATLLRLGEDDHIL--TIVMHHIISDGSVDLRRD 1251
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2067 TEMGRMAESTAHAFKLEDELPPIHVRLYQVVRDORTLSFASIVCHHLAFDAMSVDVFD 2126
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1252 LNOYLSAALKSDKPLSALPLPIQYSDFA-KWQKQDFIEQEKOL-NYWKQKQKSDSPAK 1309
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2127 LDIFYAVHTKHK--AAANLPTLRQYKEYAIEHRRALRAEQHRLADYWLRLKLSMEASY 2184
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1310 IPTDFARPALLSGDAGCVHVTIDGELYQSIRAFECNEHTTSFVYLLAAFRAAHYRLTAVE 1369
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2185 LVPDRPAPQFDYNGNDQLQSTTPTTAQLKELAKREGSSLYTVVAAAYFLLLYVYNQR 2244
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1370 DAVIGTPIANRPELEDIIGCFVNTOCMRINIDHDTFTGLTINOVKATTTAAFENEDIP 1429
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2245 DITIGIPVARNHPDFESVVGFFVNLPLRVNYSQSDIHG-LIQAVOKELVDAQIHODLP 2303
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1430 FERVVSALQCGSRDLSTPTLAQLIF-----AVHSQKDLGRFKFQGLSVPPSKAYTR 1482
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2304 FOETTKLLHW-QHDPSSRHPQLQAVFNWENPANVHEBQLQYK-----PPSLPSAA--K 2356
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1483 FMEFHLFOETDSILKGSVNFADLFFKMETVENVRVFEILLRN-GLOSSRTPVSIILPTD 1541
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2357 FDLNVTVKESVNSLVNMFNPTSLFEETVQGMETFFHLLRLLOLAHNKASTLSKLSVED 2416
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1542 GIYVLEKLDVNLKXVDYPRESSIADVFQVQSAVPSLAVDSVSSCLRTYTEL-DROSDI 1600
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2417 GVLNPEP---TNLPSSRDSGNSLHGLFEDIVASTPDRIADGTRLSYSELNERANQL 2473
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1601 LAGWLRSSMPAETLVAVFAPRCSETIVAFVGVKLANLAYLPDLDVSPSARVODIISGLS 1660
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2474 VHLIISASIVADDRIALLDKSIDMVIALLAYWKAGAAVPLDPTYPQSORTLILEESS 2533
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1661 GPTIVLIGHDTAPDIEVTNVEFVRI--RDALNDSNADGEVIEHDSKPSATSLAYVLY 1718
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2534 ARTLIITRKIT-PRGTVANVPVSVLDSPTLACLNOQSKENPTTSTQKPS--DLAYVIF 2590
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1719 TSGSTGRPKGMIEHVRIVRTVSGCIPNYPSETRMAH-----MATIAFGASYEIIYSALL 1774
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2591 TSCTTGKPKGVLEHQSUVQ-LRNSLIERYFGETNGSHAVLFLSNVYVDFSLQCLSLVL 2649
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1775 FGTILVCVDMYTTILDARALKDVFREHVNAAHVTSSSQDVPL-RVPRRLSRLTMEFFLV 1833
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2650 GGNKLI-IPPEGLTHEAFYDIGRREKLSYLSGTPSVLQQLIELSLRP-----LHMVTA 2703
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1834 VTDSTAPDALDAQGLQGVQYNGYGPTEGVNMSTIYPTIDSTESFINGVPIGRAL---N 1889
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2704 GEEFHASQEKMSQFAG-QINNAYGITE-----TTVYNIITT--FKGDAPFTKALCHGIP 2756
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1890 NSGAYVVDPEQQLVIGVGMELVVTGDGLARGY-SDKALDENRFV-----HITVNDQT 1941
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2757 GSHVYVNLRLQVPENAVAGELYLGGDCLARGYLNQDALTNERFIPNFPYEPKQASDRP 2816
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1942 VKAYRTGDRVRYRIGDGLTEFFORMTQPKIRGNRIESAEIEAALLRDSVRDAVVLQ 2001
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2817 QRLYTKGDLVRRF-RPHHLEYLGRKQDQVKLGRFRIELSEVRDAVLAISAVKEAAVIPKY 2875
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2002 NED 2004
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2876 DED 2878

```

RESULT 7

US-09-134-001C-3464
: Sequence 3464, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

```

: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134.001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3464
: LENGTH: 2404
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-3464

```

```

Query Match      8.8%  Score 1425.5;  DB 4;  Length 2404;
Best Local Similarity 23.2%  Pred. No. 1.7e-115;
Matches 582;  Conservative 466;  Mismatches 1032;  Indels 431;  Gaps 96;

Qy 68 QSAIGHAV---YDVPTDIDISRFALAWKEIVNQTPALRAFAFTSDSGKTSQVILKDSFVF 124
Db 36 QSSSEHRTVYWKLSRDIELRLMYALLDVVQSQPLRT-QFVTTDDFNOLKINLRDFF-- 92
Qy 125 SWNCWSSSSPDEVVRDEAAAAAGPRCNRFLLEDM--QTKKCOLVVTESHALVDVTFQ 182
Db 93 -----PFIEIKVEMNSQSIDLEAFFTRNLNSYHFN 123
Qy 183 QRVLSRVFAAYKHEK-----DTHRPETPESSDATDTSQSVSVVSMSCEDNAYSATHFW 236
Db 124 QLPFN-FKIYQFLDEAYLLLDLPHATIFNESQLTPFLOQLNIATYHLSKSEYSISDFYNW 182
Qy 237 QTHLN---DLNASVFPHLSHLMVNP-----TTTAHRITFPLSOKALSNSAIC 283
Db 183 IKEMNKMDQNVQVCP--SRHFNVLNADGONYAYIPVKNTECKKKMCSLHAE-LPSLDID 239
Qy 284 RTALSILLSRY--THSDEALFGAVTEQSLPDKHYLDG-----TYQTVAPLRVHCQ 333
Db 240 VWIYSIYLAHHFISQSSDVTLGI-----HFSIDNKNTENMMVNLNTDIAPLNLIS 289
Qy 334 SNLRASDVMDAISYDDRLGLHAPFLGRIDIRNTGDN-----GSAACDFQTVLLVTDGSHV 388
Db 290 QSDVWDMVDECSAL-----LEELQMGASFPVQPKAVQIDVEIMI-----HI 332
Qy 389 NNGINGFLOQITESSHPMPCNNRALLLHCOMESSGALLVAYYH-----NVIDS 437
Db 333 EK-----VQSQFELNHI--CHHIFRLYN---EASSFADLEFYPHVQDGFIVVNDNYDD 382
Qy 438 LQTRLLQOF-GHLIKLQSLPDLSSMAEYNLMTEDYDRAETESMNSQPLVQDTLIHEM 496
Db 383 LTVHTLVKLINGIYMQITQNPSSL--IKDKLSRSLAKYNDINLONNDINSEVITYKT 440
Qy 497 L-----KAVSHSPTKTAIQAWDGDWTYSELDNYSSRLAVHIKSLGLRAQQAIIIPVYPEKS 551
Db 441 VVERFERQVHOHPDSIALQVEQSRSMYVHQLNQCANLLAYRLNLHQHTEPNDMVALIAERS 500
Qy 552 KWTIASMLAVLKSGNAFTLIDPNDPPARTAQVVTQTRATVALTSKLHRETQVQLVGRCVV 611
Db 501 LEMIICMLGTLKAGAGYIPIDPDYPERMNYIITEDAKPRAVVT---YRTSFQSLGPO--- 554
Qy 612 VDDELQSVSASDDFSS---LTKSQDLAVYVFTSGSGDPKGMIEHRAESSCALKFGAS 668
Db 555 MDELIVD-SREHDINDPRGINCSEDIAVYVYSGTIGKPKGLVPHRGIDR-LVHNPY 612
Qy 669 LGINDSTRALQFQTHAFGACLLLEIMTTLINGGCVCIPSDDRMNS--IPSFIRNYNNWM 726
Db 613 VELNENTVLLSGTVAFDAATFEIYGPLLNGGRVLVITSKDTLLNPLLQDQAITEKNVNTM 672
Qy 727 MATPSYMGTFSPEDVPLATL--VLVGEQMSVSNVAIWAPKL-----OLLNGYGOSES 777
Db 673 WLTSSLFNQIASERIEALESLEYLLIGGEV---LNAKWVHLLNSRECHPQIINGYGPTE 729
Qy 778 SSICFASNSTPENNNGRAV-----GAHSWIDPNDNRLVPIGAVGELVIESGCIAR 830

```


Db 730 TT--FTTFAI-POEMPSRIPIGLPIGTTTIVVMG-----DRICGVGVGELCIGGAGLAK 783
QY 831 DYIVPPPE-----KSPFTDIPSWYPANTPPDCAKLYRTDGLARYASDGSIVCLGRIDS 885
Db 784 GYLAOPKLTAEFRFQSPENNM-----LYRSGDLVRLQEDGVIDYISRIDK 829
QY 886 QVKGIRQVELGATETHLRQOMPDDLTIVVATKRSQSANSTSLTIAFLIGSSYGNRPSD 945
Db 830 QVKGIRFRIEISEKAL-EAIRDINKAVVIREQD--DKQIVAYEASOL-----878
QY 946 AHILDHDAKAINIKLEOVLRPHSPFICYMELPRATATCKIDRRRLRIMGKIDLDKOT 1005
Db 879 -----KSTGOLKDLSETLPEYMPVHFVKVDRIPITMNGKLDVRL-----PEINLANN 928
QY 1006 QCAIVQOAPAPIPVFPADTAADKHSIWQSLGIDPATVNVGATFFELGNSITAKMVM-M 1064
Db 929 RNYVEPR-----NDIERTVCRIFEILHVD--QVGVKDNFFELGHSRLRATLVNRI 978
QY 1065 ARSVMGDLKVNIVQHPTLAGISAVV---KGDPLSYTLIPKSTHEGPPVEQSYSGRLWFL 1121
Db 979 EERLKKRLKVGDLKMSPTVEQLGQOIEELQND--VYEVIPKANESYQYDLASOKSMYLL 1036
QY 1122 DQDVGSLWYILIPVAVMRGPNVDALRRALAALEORHETLRTTFEDODGVGVQIVHEKL 1181
Db 1037 WKVNPKDTVYNIPFLWRLSSELNVQLORALKLIERHEILRTQVIDDNEVKORIATHV 1096
QY 1182 SEEMKVIDLGSDDLDPFVLNQEQTPFNLSSEAGWRATLLRLGEDDHLITVMMHIISD 1241
Db 1097 SPDPEEVT---TSLTNEQDIIQSEMEFPDLQPSQMRVKYIHGPOQDY-LFMDTHHSIND 1152
QY 1242 GWSIDVLRDLNOLYSAAKSDKPLSALPLTPQYSDFAKWKQDQFTEQEKOLNWKYKQ 1301
Db 1153 GMSNTILSLDNALYQD-----KSLPELKQYKDYSEWVHR--DLSKQRHFVLOQ 1201
QY 1302 LKDSPP-AKIPDARPALLSGDAGCVHVTIDGELYQSLRAFCEHNTTSFVLLAAPRA 1360
Db 1202 FENQVPLNMTDTPRPSIKITNGMLTFHNRQIKQOLKSVIYEQHVQVDFWFASALMV 1261
QY 1361 AHYRLTAVEDAVIGTPIANRPELEDIIGCFVNTQCMRINIDHHDTFCTLINQYKATTT 1420
Db 1262 LLHKYTRQDDIAIGSVISARTHRTENMLGMFANTLVYRGRPHDQKTDWOLMAEMKEMCL 1321
QY 1421 AAFENEDIPFERYVALOPGSRDLSSSTPLAQLIFAV-HSQKDLGRFKFQGLSVPPVPSKA 1479
Db 1322 GAYEHQEPFESLWNL--VDERDASHNPLFDVMLVLQNNETHANFNGHSQTHIP-POST 1379
QY 1480 YTRFDMEPHLFOETSLKGSVNFADFLFKMTENVVRVFFELRNGLOSSRTPVSIPL 1539
Db 1380 TAKFDLSIIBEDQDDYVNIYNTDLYKQETHIHAEQLOMIKHVISTENLKIQDIDE 1439
QY 1540 TDGIVTLBKDLVNLVKHV-----DYPRESSLADVFQTVQVSAYPDSLAYVVDSSCRITYTEL 1594
Db 1440 NDDL--LTLWLD-----KHVNDCSLDLPKNKSIQQLLHDVYMKAKADVDALKMNGQSMYQEL 1493
QY 1595 DRQSDILAGWLRRRSPAEITLVAVPAPRSCETIVAFGVKLANLAYLPDVRSPARVQD 1654
Db 1494 DQYSNSMAQTIONGIRGERVALLTERSFEMASMIYAVLVKGSYVPIDVTPDKRIEF 1553
QY 1655 ILSGLSGGTIVLIGHDTAPPDIEVTVNVEFVRDALNDSNADGFV-----IEHDSKPS 1709
Db 1554 IIEDAEVAALVITYGKAIS-----SHIPVIEDIDNTENKRLNIEYAGNLEDD-----1602
QY 1710 ATSLAYLVLTSGSTRPKGVMEHRIIVRTVTSICI-----PNYPSETRMAHMAITAFDGA 1765
Db 1603 ----MYHIYTSQTKPKAVSKORNILNV---CAWTKRLNLSDDDEVYLOYVANYVPFAS 1655
QY 1766 SYEITYSALLFGRITVCVDYMTLDRALKDVFREHVNAAASHVTSSSQDVLVPRRLSR 1825
Db 1656 ATDFYCSLLNGPLVIATSVERTNTDLEKLIQSENITIAS-----IPLQVYN 1703
QY 1826 TLMFEFL--VYTDSTAPDALDAQGLYQGVOCY-NGYGPTENGVMSTIYPIDSTESFINGV 1882
Db 1704 VMHIFYIPKVITGGAPSTPAFVOHISKHCDMIVNAIGPSENTVITTSWIYKGDAPSTI 1763

QY 1883 FIGRALNNNGAYVVDPEQOOLVIGVGMGELVWTGDGLARGYSK-ALDENRFVHITVNDQT 1941
Db 1764 FIGKPLANVDIFIMS-GKLCGVGIPGELCIAGESLTSGLNRPPELSAEKFINNPFPG- 1821
QY 1942 VKAYRTGDRVYRIGDGLIEFGRMDTOFKRGNRIESAEAEALLROSSVDAVAVLQO 2001
Db 1822 -OLYRSGDLARL-MPDQIEFLGRIDKQKVHGYRIELGETEININSVDVTYDVSIVILAK 1879
QY 2002 NEDQAPETILGFVADHDHSENDKQSANOVQEGWQHFESGMYSDIGEIDPSDPSIGSDFKGW 2061
Db 1880 QGER--EVLHAYYVGSQDEENHISOLNAY--LPKYMIPKLTALISEI-PLT-GND-----1929
QY 2062 TSMVDCSQIDFDEMHGELGETTRTLHDNRSLGNVLNLEIGTSGMI---LFNLDLSRLSYVG 2118
Db 1930 --KVDESRLPVPNVHKNKFVAPRN-NIEREIAQIV-----SGVLDVSSMSIDD-----1974
QY 2119 LEPSRSAANFVNKATESIPSLAGKAKVQVGTATDQGVDDHLPLDLVLVNSVIQYFSPSEY 2178
Db 1975 -----DFTFMGCTSLDAMVVVSKL-----KNGIHI---ITMQDVYQ-EKTVRY 2013
QY 2179 LAE-----TADTLI--HLPNVQRIFFGDRVSQATNEHFLAARAIHTLGKNAT-----2223
Db 2014 IANTHEKROALPEVVLPDHLPOLOSIV--ERYQLKSOHLTQSSGLHVLTLGATGFLGAY 2071
QY 2224 -----KDDVRQKMAELEDKEEELVPEPAFTSLKDRFPGLVHEVILP 2266
Db 2072 LIDEMQDDAQITCIVRGHDINQAKTNLENN-----LNCYPDTAHVK---LMKHIDIIL 2123
QY 2267 KMEAVNEL---SAYRYAAVHVRSGLDELVPVLEKDDWDIDFQANQLNOKSLGOLLKSS 2323
Db 2124 ADLSLDELHIIDSA--IDTIIHA-GARTDHF-----GDETFDFVNVVRSTOALIDLANKK 2175
QY 2324 DAAIMAVSKIPPEITAFERQVVASLNSNIDEMQNSTIRSSAEGDSLSVPDIFRIAGEAG 2383
Db 2176 KAKLIYISTI-----SVGTVEVHQDDITTFSEKDIYK--GOL- 2210
QY 2384 FRVESSARQWSQNGALDAVHHCCSQGRTLVNF--PTDHHLRGSDLLTNR 2432
Db 2211 FTSPTYKSKFYSEIKVLEAVNEGLAAQIIRLGNLTASGTGPLNMKNLTNR 2261

RESULT 8

US-09-413-814-11
; Sequence 11, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCI/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2628
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-11

Query Match 8.6%; Score 1392; DB 4; Length 2628;
 Best Local Similarity 25.3%; Pred. No. 1.9e-112;
 Matches 562; Conservative 341; Mismatches 843; Indels 474; Gaps 80;

QY 1112 SYSOGRWFLDQDVGSLWILYIPYAVRMGRPNVDALRRALAALEQRHETLRRTTFEDODG 1171
 DB 24 SFAQRRLWFLDQLEPGSAVYNMPASFRTRGPDYDVSLSVRSVNEIVRHESLRITVDVIG 83
 QY 1172 VGOIVIEKLESEMKVYIDLQSGDLDPFE-----VLNQEQTTTFNLSSEAGWRATILRL 1224
 DB 84 EPVOVIAPSLRIEVPVVDL--SEIDEPEREAEARLMAEESRRPFDLTRGPLLRAKLLRL 141
 QY 1225 GEDDHITIVMHHIISDWSIDVRLDNLNOLYSAAKDSKDPALSALPLPLOTYSDFAKW 1284
 DB 142 GEADHVLLIWMHHIVSGWMDVLFKELSTLYAAFHGRPSP---LEPLPIQYADFVWQ 198
 QY 1285 KD--QFTEQKQLNWKYKQKLDK--SPAKIPTDFARPALLSGDAGCVHVTIDGELYQSLRA 1341
 DB 199 RELQGEVLESHGLYWEHLRGAPTLLELPMDRPPPAQTRFGRQRAFRPLSLSQAVQA 258
 QY 1342 FCNEHNTTSFVWLLAARAAHYRLTAYEDAVIGTPIANRNRPELEDIIGCFVNTQCMRIN 1401
 DB 259 LSHOEGATPMTLLTAFSVLLSRVARSQDLWGTPIANRTRAELEGIGFVNMALRID 318
 QY 1402 IDHDTFTGLINOVKATTTAAFNEDIPFERVVSALQPGSRDLSSTPLAOLIEAVH---- 1457
 DB 319 LGGDPSPRELLGRVREVTLCAYAHQDLFFELVEELSPG-RSPSHSPLFQVSTLQNTPM 377
 QY 1458 ---SQDKLGRKFGQLESVPVSKAYTRFDMEFHQLFQETDSLKSGVNFADLEKMETVEN 1514
 DB 378 DATNRADIAS-----GGAPLVEKMAAKFDLILELSESPQGLLTPEYNTDLDAGTIER 431
 QY 1515 VVRVFFELLRLNGQSSRTPSYILPLTDGIVTLEKLDVNLVHVDPRESSLADVFQTVS 1574
 DB 432 MAGHLEVLLSSAAVADPRPAELPLMGAERSRVLVENWSTAALYPEDHCHMHELFEOQVE 491
 QY 1575 AYPDSLAVDSSCLRTITELDRQSDILAGWLRRSRMPAETLVAVFAPRSCETIVAFRGVL 1634
 DB 492 RSPENTAVLQOQTLTYRELNMARANLAHLRSLGVGPVVRVGLYLERSTETVAILGVL 551
 QY 1635 KANLAYPLDVRSPSARQDILSGLSGPTVLIGH---DFAPPDIEVNTVYFVRIRDALN 1691
 DB 552 KAGGAYVPLDPTYPEERL-GLNMADAAPSLLTQASLLSKLPPHGDATLVOLDALHEALS 610
 QY 1692 DSNADGFEVTEHDSKTPSAT--SLAYVLYTSGTGRPKGMVIEHRYI--IRTVTSGCIPN 1747
 DB 611 R-----LPHTPRSGVTAQNLAVMYTSGTGRPKGVLEHRLGCLNPTVOAKLYGI 662
 QY 1748 YPSETRMAHMATIAFDGASYEISALIFGRITLVCVDYMTLDRALKDVFREHVNA--- 1804
 DB 663 APGD-RLLQFAPLCPDTSFCEIALALLSGATLVMTAGDELPGPLVELLKKHATVAMLL 721
 QY 1805 ---ASHVTSSQDVPVRRLRSLTLPFFLVITDSTADALDAQGLYQGVQVNGYG 1859
 DB 722 APTVLAALPEQQAALPLRY-----LTMAGEACPABELVRWRKAPGRRLFNSYG 769
 QY 1860 PTENGWMTYIPIDSTESFINGVPIGRALNNSAYVVDPEQQLVGVGVNGELVVTGDGLA 1919
 DB 770 PTETTWASADLSERI---PPIGRPANTQIYVLDLEALEVPVIGVFEIPFGVGVA 826
 QY 1920 RGYSDKA-LDENRFVHITVND-QTVKAYRTGDRVRYRIGDGLIEFFGRMDTQFKIRGNRI 1977
 DB 827 RGYHGRPDLTAERFVDPFGQTKGARLYRTGDRARW-LPDGNLEFLGRNDEQVKVRGRI 885
 QY 1978 ESAEIEAALLRSSVDAAVVLOQNEQAPILGFVVDHDSKNDKGSANOVEGWQDH 2037
 DB 886 ELEEIEAALLKHPAQAQAVAVVREDTPPGDKRLVAVV-----GRGARVTA--- 931
 QY 2038 FTSGMYSDIGEIDPSTIGSDFKGTWYDSQIDFDEMHWEHLCETRTLHNRSLGNVLE 2097
 DB 932 --AELRQVSERLIPAT-----WVPSFVALDA-----LPLTPNGKVDRLALP----- 971
 QY 2098 ICTGSGMILNLDLSRLSYVGLSPRSAAAFVNKATESIPSLAGRAKVQVGTATDIGQVD 2157

DB 972 -----EPEQSA-----GCGD 981
 QY 2158 DLHPDLVNLNSVIOYFSPSEYLAEI-----ADTLHLPNVORIFFGDVR----- 2201
 DB 982 HVAPRNAVBEELARIWASVLRKRGVHDNFEIGGDSILSIQIVVRAAQAGRLRTPROM 1041
 QY 2202 --SOATNEHFLAARAIHTLGNATKDDV-----ROKMAELEDMEBELLV-- 2243
 DB 1042 FOHQTI AELSTVARAVEAV--HVQDPVTPGAPLTPVORWLEQEAEPHFHNSIFLEV 1099
 QY 2244 -EPAFFTSLKDRFPGLVHVEI-----LPKNMEAVNELSAYRYAAVHVHVGSLGDELVLV 2298
 DB 1100 RERDESALQOATAHLJDHDLRLRLARDERGAHVFAAPGGSTPQRVDLG---ALP- 1155
 QY 2299 EKDDWIDFQANQLN--OKSLGDLKSSD-AAIMAVSKIPPEI--TAFEROVVASINSNID 2353
 DB 1156 -----SAEQISAMEKAASEAQASLDLAAGPVVRVAVLDGGEVAPQRLLVIAHHAIVD 1207
 QY 2354 --EWOLSTIRSSABGDSLSVPDIFRI-----AGEA-GFRVEVSSARQW-----SONGALD 2401
 DB 1208 SVSWR-----ILLDOLFGAYEQARRGEAVRLPPKTTSVKRWAEELLTEHAGSE 1254
 QY 2402 AV---FHHCCSOGRTLVNFTDHHLRGSDLLTNR-----PLQRLQNRRTAIEVRRLRS 2452
 DB 1255 AVRAELGYWLDSSRRTVAPLPVDRR-AGEDVWGSARHIVVSLTPEQTEQLLREYPOAYRT 1313
 QY 2453 LLPSYMIIP-----SNIVVLDKMLNANG-----KV 2477
 DB 1314 RIDDALLTAFAATARTWGTSPAVLLD---LEGHREELAGVLDLRTVGVTFATMYPILLRV 1370
 QY 2478 DRKELSRRAKVPKQQAAP-----LPTFPISEVEVI-----LCBEATEVFGMKV 2522
 DB 1371 DAADPGEALKSIKEQLRAVPGRGLYGLLRYSRDTIAEVRALPQAEIC----- 1419
 QY 2523 DITDHFENLGGHSLATKLISRIDORLKVITVKNDFVDFADLASVIRQGLGLQLOQVPS 2582
 DB 1420 -----FN-----YLGQDQAI-----PEAAPF-----RPAR 1440
 QY 2583 DGOQDORS--AHMAPRTETEAILCDEFKVLGVQVGTITDNFFDLGGHSLMATKLAVRIGH 2640
 DB 1441 EYQGSERSPGAHRAHLIEVNA-----SIANGRLVATWTYSRRH 1479
 QY 2641 RLDTTVSVKVDVDFHPVLFQALAIADNLVQSKTNEIVGREGMAEYSPQLLFTDEPEFMA 2700
 DB 1480 EPETIERVAASF-----VTALRALIAHCTLPFEGGNTPPSDFKVLRL-QETIDALDA 1530
 QY 2701 SEIKPOLEIQ-ETIODIYPTSTOMKAFLDHTTARPPFVPEYIDPSTSEP--DAAGLI 2757
 DB 1531 IDAGPSPSARGSRIEDVYPLSPQEGILFH--TLYATDYATYVEQFHWTLLEGDFDAEFT 1588
 QY 2758 KACESLVNHLIDIFRTVFAEASGELYQVLSCLDLPQIVET-----EDNINTATNEF 2809
 DB 1589 RALQDVVARHAALRTSFA-----WERLDAPLIQVTRGAVLPVEHQDLRLAAEEQ 1638
 QY 2810 LDEFK--EPVRLGH-----PLIRETIKQTK-SMRVIMRISHALYDGLSLSEHVVKLH 2860
 DB 1639 TAHISRYVEAERQRRDLRKAPLMRAGLLRLRDAWCLVETIHHLLIDGWSTQILLKEVF 1698
 QY 2861 MLYNGR-----SLLPHQFSRYMOKTA-DGRESHGCFWRDVTION---TPM---T 2902
 DB 1699 TLYEAHRGHRGHLEALEQPRPYGDIYIGLAKQDQVTRTAARRELEGFSAPPLGVDR 1758
 QY 2903 ILSDO-----TVVGDNDATCKALHLSKIVNIPSOVLGRSSNIITQATVFNAAACALV 2953
 DB 1759 VPDDGPRGFRGRRIRALSQDGAARLAAE-----ARQHOLTSTLTVQGAALL 1805
 QY 2954 LSRSDSKDVGRIVTSGROGLPVEYQDVGPCPNVAVRAHIESDYNOQ---LHDIDQD 3011
 DB 1806 LSRYSQDPOVLFGMTVSGRSAPITGPIERTGLFINTIPVRVR-EPADASYLAWLALQEH 1864
 QY 3012 YLLSLPHETIGFSDLRNCTDWPEAITNPSCCITTYHNFYHPESQFE-----QORVEMG 3065

| | | | |
|----|------|--|------|
| Qy | 1565 | LADVFQTVQVSAYPDSLAVVDSSCRLLTYTELDQROSDILAGLRRSMP-AETLVAVFAPRS | 1623 |
| Db | 1419 | LHEMFENASQPKIAVVIYEETSLTRELNERANMAHQLRSDVSPNPNIEVALVMDKS | 1478 |
| Qy | 1624 | CETIAFFGVLKANLAIPLDVRSPSARVODILSGLSPTIVLIGHDTAPDPEVTNVF | 1683 |
| Db | 1479 | EHMIVNILAVWKSAGVPIPDGYPNDRIQVILEDTQ-ALAVIADSCYLPRIKGMASG | 1536 |
| Qy | 1684 | VRIDALNDSNADGEFVIEHDSKPSATSLAYVLYTSGTGPRKGVGMIEHRVITRTVSG | 1743 |
| Db | 1537 | TLLYPSVLPANPDSKWSVSNPSPLSRSTDLAYIIYTSCTGRPKGVTVHHGVVQLSV | 1596 |
| Qy | 1744 | C---IPNYPSETMAHWATIAPDGASEIYSALLFCRTILVCV-DYMTTILDAALKDVF | 1798 |
| Db | 1597 | SKVGLKDTDEVILS-FSNYVDFHFVEQMTDAILNGOTLLVNDGM-----RGDKERLY | 1650 |
| Qy | 1799 | RE-HVNAASHVTSQQDVLVRPRLSRITLMMFFLVVTDSTAPDALDAQGLYGVOCYNG | 1857 |
| Db | 1651 | RYIEKRNVTYLSGTPSVSVSVESRPKDLRLRVDCVGEAFSEPVFKIRFTFHLG-VING | 1709 |
| Qy | 1958 | YGPTENGVMS--TIYPIDSTESFINGVPIGRALNNSGAYVDPQOVLGVGMELVVTG | 1915 |
| Db | 1710 | YGPTEVSIITHKRLYPPPPERRM--DKSIGQGVHNSYVLNEDMKRTPIGAVGELYLG | 1766 |
| Qy | 1916 | DGLARGYSKKA-LDENREFHITVNDQTVK-----AYRTGDRVRVRIIG-DGLIEFFGRM | 1966 |
| Db | 1767 | EGVVRGVHNRADVTAERFIPIPNFOSEDKREGNSRLYKTGDLVRWIPGSSGEVEYLGN | 1826 |
| Qy | 1967 | DTQFKIRGNIESAEIAALLRDSVRDAAVLQONEDQAPAIL-GFVAV-----2016 | |
| Db | 1827 | DFQVKIRGLRIEVEIEAILSSYHGIKQSVVIKADCRGAKQFLVGYVADAALPSAAIR | 1886 |
| Qy | 2017 | -----HDHSENDKQGSANOVGE-----2033 | |
| Db | 1887 | RFMQSLRPGYMWPSRLILVSKFPVTPSGKLDTKALPPAESEIDVPPRSEIERSLCDI | 1946 |
| Qy | 2034 | WQDHFS-----GWYS-----2044 | |
| Db | 1947 | WAELEWHPEIGIYSDFFSLGGSLSKTKLSPIMHESFNRAVSVSALFCHRTVEAQTHL | 2006 |
| Qy | 2045 | -----DIGIEDSTIGS-----DFKGWTSYMDGSDQDFDEMHEWLGE | 2081 |
| Db | 2007 | ILNDAADVHEITPIDCNDQMIPVSAQERLLFIHEFENGSNAY-----NIDAAFELPGS | 2061 |
| Qy | 2082 | TTTTLHDNRSLGN-----VLEIGTSGMILFNLOS-----2111 | |
| Db | 2062 | VDASLLEQALRGNLARHEALRTLVLKDHATGIYLRQVLSPDQAQGMFSNVNVDATAKOVERL | 2121 |
| Qy | 2112 | -----RLES-----YV | 2117 |
| Db | 2122 | DQBIASLSQHVFLDDELPEARILKESGGLYLILAFHHTCFDAMSLKVFOELRALYA | 2181 |
| Qy | 2118 | GLEPRSAA-----2126 | |
| Db | 2182 | ALQKTKSAANLPALKAQYKEYALYHRQLSGDRMRNLSDFWLKRKLGLEPLQLITDRPP | 2241 |
| Qy | 2127 | -----AFVNMKATESIPSLAGAKVQ-----VGTATDIGQVD-----2157 | |
| Db | 2242 | VQFKYDGDLSIELSKKETENLRGVAKRCKSSLYVVLVSVYCVMLASVYANQSDSVGIPV | 2301 |
| Qy | 2158 | --DLHPDLVNLVNSVIOYFPS-----SEYLAETADTLIHL-----2189 | |
| Db | 2302 | SHRTHPQ---FQSVIGFFVNLVLRVDISQSAICGLIRVMKELVDAQLHODMPFOEVTK | 2358 |
| Qy | 2190 | -----PNVQRIFFGVRSDQATNEHFLAARAHTLGN-----ATKDDVR | 2228 |
| Db | 2359 | LLQVNDPDRHPLVQNVF--NFEESRANGEHDAARSEGSLAFNQRPVQPVDSVAKFDLN | 2416 |
| Qy | 2229 | QKMAELEDMEELLVEPAFTSL--KDRPFGVLVEHVEILPKNM-----EAVNE--LSAY | 2278 |
| Db | 2417 | ATVTEL---ESGLRVNFNATLSFNKSTIQGFHTVYELLRSLSELSAEGINDELQLSLV | 2473 |

| | | | | | | |
|----|------|--|--|-----------------------|-------------|------|
| QY | 2279 | RYA | -AVVHV----- | RGSLGDELVLVPEKDDWIDF | 2300 | |
| Db | 2474 | RPTENGDLHLPLAQSLPATTAAEOKVASLNOAFEREAFLAAEKIATVVOGDRALSYADLNG | 2533 | | | |
| QY | 2307 | QANOL----- | NOKSLGD----- | LLKSSDA--AIMAVSK----- | IPFEIT----- | 2338 |
| Db | 2534 | QANOLARYIQSVSIGIGADDGIALMLEKSIDITICILAIWKAGAAAYPLDPTYPGRRVOL | 2593 | | | |
| QY | 2339 | --AFERQVVASLNSNIDEWQ----- | LSTIRSAEGDSLSLPDIFRIAGEAGFRV | 2386 | | |
| Db | 2594 | LEEIKAKAVLVHSSHASKCERHAKGAVTADSPALETAVSQSOAADLPTIASLGNLA--YI | 2651 | | | |
| QY | 2387 | EVSARQWSONGAL----- | DAVFHH----- | C | 2407 | |
| Db | 2652 | IFTSGTSGKPGKVLEQKAVALLRDALURERYFGDCTKHHGVFLSNVYDFDSVEQLVLS | 2711 | | | |
| QY | 2408 | CSQORTLVNFTD----- | HHLRGS--DLLNRPRLQRLQNRRTAI-- | 2444 | | |
| Db | 2712 | VLSGHKLIVPAEPVADDEFYRMASHGLSYLSGTPSLLOKIDLARDHDLQVVTAAGEEL | 2771 | | | |
| QY | 2445 | -----EVRER----- | LRSLLP----- | SYMPSNI | 2463 | |
| Db | 2772 | HATQYKMRFRFNGPIYNAYGVTEVVYIIAEFTTNSIFENALREVLPCTRAYVLTA-- | 2829 | | | |
| QY | 2464 | VLDKMLPLNANGKVDKRELSRAKVPKQATAPLT----- | FPISEVEVILCEATEV | 2517 | | |
| Db | 2830 | -ALQVPFVDAG----- | ELYLAGDTVTGRVLYNQPLLTQDORFIPNPCKEEDIAMGRFARL | 2883 | | |
| QY | 2518 | F----- | GMKVDTID----- | 2526 | | |
| Db | 2884 | YKTCGLVRSFRNQOQPOLEYLGRGDLOIKMRGYRIEISEVQNVLTSFGVREGAVVAKY | 2943 | | | |
| QY | 2527 | ---HFFNLGGHSLA----- | TKLSRID--QRLKVRITVKDVFDB----- | PVFADLA | 2568 | |
| Db | 2944 | ENNDYTSRTAHSLVGVYTTDNETVSEADILTFMKARLPTVWPVSHLCCLEGALPVTINGK | 3003 | | | |
| QY | 2569 | SVTROGLGLQOVPYSDGQGDORSAAHMAPRTETAILCDEFAKVLGFQ--VGITDNFDFLGGH | 2627 | | | |
| Db | 3004 | LDVHR---LPEIINDSA-- | OSSYSPPRNIIKAMCRLWESALGMERCIGIDDLFKLGGD | 3057 | | |
| QY | 2628 | SLMATKLAVRIGHRLDPTVSVKDVFDPHPLFQALIALDNLVQSKTNEIVGGREMAEYSPF | 2687 | | | |
| Db | 3058 | SITSLHLVAIHNQVCGKITVRDFEH----- | RTARALHDHVFMKDS----- | DRSNVTQF | 3107 | |
| QY | 2688 | QLLFTEDEPEFMASEIKPOLEQEIITODIYPSTOMOKAFLEFHTTARPPFPVFDIDPPS | 2747 | | | |
| Db | 3108 | R---TEOGPVIGEAPLIP----- | IQDMFLSKALQHPMYNHT----- | FYV---R | 3145 | |
| QY | 2748 | TSPDPAAGLLKACESLVNHLNIDIFR--TVFAEAGELYQVVLSCLDLPQVLETED--NINTA | 2805 | | | |
| Db | 3146 | TPELDVDLSAAVRDLDQHDHVFMRKLKREEVGFVQSFADFSFPAQLRVNLNVKDDGGAA | 3205 | | | |
| QY | 2806 | TNFEFD----- | EFAKEPVRGLHPLIRFTIITIKOTKMRVIMRISHALYDGLSLSHVVRKLUHM | 2861 | | |
| Db | 3206 | VNEILDGWSGFNLNGPIGS-- | IGYLHGVEDRSARVWFESHVHMAIDTVSMQILYRDLQT | 3263 | | |
| QY | 2862 | LYNGRSLLP----- | HQFSRYMQ--YTADGRESCHGFWRVIOQNPMTIILSDDTTVDVDCNDAT | 2916 | | |
| Db | 3264 | LYRNGSLGSGKGSFROWAEAIQNYKASDSERNH-- | WNKLVMETASSISALPT----- | ST | 3315 | |
| QY | 2917 | CKALHLSKWINPISQVLRGSSNIITQATVFENACALVLSRESDSKDV--VEGRIVSGROGL | 2975 | | | |
| Db | 3316 | GSVRLSRLS--PEKT----- | ASLIOGGIDRDQVSVYDSLLITS--VGL | 3355 | | |
| QY | 2976 | PVEYQDIVGCTNAVVPRAH | 2995 | | | |
| Db | 3356 | ALQHIAPTGP-- | SMWTIEGH | 3373 | | |

RESULT 10
US-08-222-617A-2
; Sequence 2, Application US/08222617A
; Patent No. 5882879

GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liemert, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3778 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-222-617A-2

Query Match 8.3%; Score 1342.5; DB 2; Length 3778;
Best Local Similarity 21.2%; Pred. No. 9.3e-108;
Matches 730; Conservative 515; Mismatches 1232; Indels 963; Gaps 134;

QY 312 FKHYL-----ADGTYQTAPLRVHCOSNLRASD-----VMDATSSYDDRLGHLAPFGLRDI 363
DB 141 FSVHQLKGFNGTHTITASL--HREQNLQNSPSSWVSPTIVTHENRDGWSVAQAVESI 198
QY 364 ---RNTGNGSACDFOQVLLVTDGSHVANGINGFLQQTIESHF-MPCNNRALLLHCQM 419
DB 199 EAARGEKESVTAID-----SASSLVKMGFLDLLVSFVDADARIPCFDFPL-----245
QY 420 ESSGALLVAYYDHNV-----IDSLQTRTLQOFGHLIKCLOSPLE---DLSSMAEYVNLMT 470
DB 246 ----AVIVRECDANLSLTFSDCLENEETICNFTDALNLLAEAVIGRVTPVADIELLS 301
QY 471 EYDRATESNWSOPLVQDT-LIHHEMLKAVSHSPKTAIQAWGDWTVSELDNVSRLA 529
DB 302 AEQKQOLEEWNNTDGEYPSKRLHLHIEVEVHERHDKIAVVCDERELTYGELNAQNSLA 361
QY 530 VHKSJGLRAQQAIIIPVPEKSKWVIASMLAVLKSGNAFTLIDPNPPARTAAQVVTQTRA 589
DB 362 RYLRSGTILPEQ-LVALFLDKSEKLITVILGVKSGAAAYVPIIDPTYDERVRFVLDTKA 420
QY 590 TVALTSKLHRETQVKLV----GRCVVVDDELQSVASD-----DFSSLTKSODLAY 637
DB 421 RAIASNQHVRLQREIVIGDRNICIIRLEPLASLAQDSKFFPAHNLDDLPLT-SQOLAY 479
QY 638 VIFTSGTGPDKGIMIEHRAFSKALCFKFGASLGINSDDRTRALQFGTH-----AFGACILLE- 691
DB 480 VVTYSGTGPFGKIFKQHTNVNVSITDLSARYGV-----AGQHEAILLFSACVFEP 531
QY 692 ----IMTTLINGCVCIPSDDDR--MNSIPSFNRYNVNMNMTATPMTGTFSPEDVFGLA 745

DB 532 FVRQTLMAVNGHLLAVINDVEKYDADTLLPFTLRHSITYLNGTASVLOQYDFSDCPSLN 591
QY 746 TIVLVGEOMSSSVNAIWAP--KIQLLNGYGOSESSSI---CFASNKSTPNNMGRAV-G 798
DB 592 RIILVGENTEARYLALRORFKNRILNEYGFTSFVAFVALKIFDPSTRKDTSLGRPVNR 651
QY 799 AHSWVIDPNDINRLVPIGAVGELVIESPGIARDYIVPP---PPEK---SPFDTIPSWYPA 853
DB 652 VKCYILNPS--LKRVPIGATGELHIGLGISKYGLNRPETLPHREIPNPPOTDCEKQLGI 709
QY 854 NTFPDCAKLYRTGDARYASDGSIVCLGRIDSQVKIRGQORVELGATETHLRQOMPDLTI 913
DB 710 NSL-----MYKTGDLARWLPNGEYVLGRADFOIKLRGIRIEFGEIETMLAYPVRTSL 764
QY 914 VVEATKRQSANST--LIAFLIGSSVFGNRPNSDAHILDHDATAKAINIKLEQVLPKHSIP 971
DB 765 VVSKLRNGPEETNEHLYGYVCDSS---ASVSEADLLSF-----LEKLPYMWIP 812
QY 972 SFYICMLELPTATGIDRRRLRMKGIDLDKOTQGAIVQOAPAPIPVFADTAALKHSIW 1031
DB 813 TRLVQLSQIPVNVNGKADLRALPAVD-----ISNSTEVRSLRGDTETALGEIW 861
QY 1032 VQSLGIDPATVNVNGATFFELGGSITAIKMNWARS---VGMCLKVSNIIYQHTLAGISA 1088
DB 862 ADVLGARQSRVSRNDNMFRLGHSITCIIQIARIQRQRSLVSISVEDVFATRLERMAD 921
QY 1089 VVKGDPLSYTLIPKSTHEGP---VEOSYSQGRLWFLDQLDVGSLW-----YLIP 1134
DB 922 LLQNKQOEKCDKP---HEAPTELLEENATDNIYLANSLQOGFVHYLKSWEOSDAYVMQ 978
QY 1135 YAVMRGPVNDALRALAALAEORHETLRTFE-----DOD-----GVGV 1174
DB 979 SVLRYNTTLPDLFORAKWHAQSFALRLRFSWEKEVFOLQDPLDWRFLYFTDVA 1038
QY 1175 QIVHEKLSSEMKVIDLGGSDLPFEVNLQEQTFNLSSEAGRRATLLRLGEDDHILTIV 1234
DB 1039 GAVEDRKLEDLR-----RODLTERFKLDVGRFLVFLIKHSENRTCLFS 1083
QY 1235 MHHIISDGSIDVLRRDLNQLYSAALKDSDPLSALTPLPTQYSDFAKWOKDOFIEQEK 1294
DB 1084 CHHAILDGSLSPLLFKVVHETYLQLLHG--DNLTSMDPP-----YTRTORYLHAHREDH 1136
QY 1295 LNTWKOLKDDSPAKIPTDFARPALLS-----GDAGCV-----HVTI--DGELY-QS 1338
DB 1137 LDFWAGVQ-----KINERCDMALLNERSRYKVLADYDQVQEQRHVTIALSGDALAD 1191
QY 1339 LRAFCEHNTTSFVLLAAPRAAHYRLTAVEDAVIGTPIANRNPRL--EDIIGCFVNTQ 1396
DB 1192 LRQCSAQGITLHSIILOFVWHAVLHAYGGGTHITGITISGRNLPILGIRAVGPYINT- 1250
QY 1397 CMREINIDH----HDTFTGLINOVKATTTAAAFENEDIPFERV-VSALQPGSRD----LSST 1447
DB 1251 -LPVLVDHSTFKDKTIMEALEDVQAKVNVNMRGNVELGRHLKHTDLKHLGFLDSFLVLENY 1309
QY 1448 PLAQLIPAVHSQKDLGRKFQGLEVPVPSKATRFDMFHLFOETDSLKG---SVNFAD 1504
DB 1310 PNLDKSRTLHQTELQYSIEGGTEKLNYPILAVIAR-----EVETTGFTVSYCIYAS 1360
QY 1505 ELFKMETVENNVVFFELRLNGLOQSSRTPVSIPLTDGIVTLEKLDVLNVKHVDYDPRESS 1564
DB 1361 ELFEEMWISSELLHMVQDTLMQVARGLNEPVGSLEYLSSI-OLEOLAANWATEAEFF-DTT 1418
QY 1565 LADVFOQVSAYPDSLAVVDSRCLTYTELDROSDILAGLWRRRSMF-AETLVAVFAPRS 1623
DB 1419 LHEMFENEASQPKTAVVYEETSLTYRELNERANRMAHQLRSDVSPNNEVALVMDKS 1478
QY 1624 CETIVAFFGVYKANLAYPLDVRSPSARVQDILSGLSGPTIVLIGHDTPADPIEVNTVEF 1683
DB 1479 EHMIVNILAVKSGGAYVPIDPGYPNDRIQYILEDTQ--ALAVIADSCYLPRIKMGMAAG 1536
QY 1684 VRTDALDNSNADGFEVIEHDHKPSATSLAYVLTSGTGRPKGWIHRVIRVITGSC 1743

1537 TLLYPSVLPANPDSKWSNSPISLRSSTDLAYIIYTSOTGRPKGVTEHHGVNVLQVSL 1596
1744 C-----IPNYPSETRMAHMAITAFDCASYEIIYALLFGRTLCVC-DYMTTLDARALKDVFF 1798
1597 SKVGLRDTDDDEVILS-PSNVFDFHEVQMTDAILNGOTLLVNDGM-----RGDKERLY 1650
1799 RE-HVNAASHVTSQDQVPLVRPRLSRTLMFFFLVTDSTAPDALLDAQGLYQGVQVNG 1857
1651 RYIEKNRVYLSGTPSVVSMYEFGRFKDHLRVCVGEAFSEPVDFKIRETFHGL-VING 1709
1858 YGPTENGWMS--TIYPIDSTESFINGVPIGRALNAGYVVDPEQQLVGIGVMGELVVTG 1915
1710 YGPTESVITHKRLYPPEPRM---DKSIGOOVHNSYSYVLNEDMKRTPIGAVGELYLG 1766
1916 DGLARGYSKA-LDENRFVHITVNDQTVK-----AYRTGDRVYRIG-DGLIEFFGRM 1966
1767 EGVVGYHNRADVTAERFIPNPFQSEEDKREGRNSRLYKTKDLYRWIPGSSGEVEYLGRN 1826
1967 DTQFKIRGNRIESAEIAALLRDSVSDRAAVVLOONEDQPEIL-GFVVAD----- 2016
1827 DFQVKIRGLRIEVEGEIEAILSSYHGIKQSVVIAKDCREGAKFLVGYVADAAALPSAAIR 1886
2017 -----HOHSENDKQOSANOVEG----- 2033
1887 RFMQSRLPGYMPVSRLLVSKFPVTPSGKLDTKALPAAEESIDVVPVPRSEIERSLCDI 1946
2034 WOHFES-----GWYS----- 2044
1947 WAELEHMPBEGIGYISDFSSGLGDSLSKLSFMIHESFNRAVSVALFCHRTVEAQTHL 2006
2045 -----DIGETIDPSTIGS-----DFKGWTSMDYDGSQIDFDEMHEWLGE 2081
2007 ILNDAADVHEITPIDCNDTQMPVSRQAQERLLFHEFENGSNAY-----NIDAAFELPGS 2061
2082 TTRTLNDRNSIGN-----VLEICTGSGMILFNLD----- 2111
2062 VDSALLEQALGNLARHEALRTLVLKDHATGIYKQLVSPDEAQGMFVNVVDTAKQVERL 2121
2112 -----RLES----- 2117
2122 DQETASLSQHVFRLLDELPEARLTKLESGLYLILAFHHTCFDAMSLKVFQELRALYA 2181
2118 GLEPSRSAA----- 2126
2182 ALQTKSAANLPALKAKQKEYALYHRRQLSGDRMRNLSDFWLKLGLEPLQLITDRPRP 2241
2127 -----AFVNKATESIPSLAGKAKVQ-----VGTATDIDQVD----- 2157
2242 VQFKYDGGDLSIELSKKETENLRGVAKRCKSSLYVVLVSVVCMVLAANYQNSDVSVGIPV 2301
2158 --DLHPDLVLVNSVIOYFPS-----SEYLAETADTLIHL----- 2189
2302 SHRTHPO---FQSVIGFFVNLVLRVDISQSAICGLIRVMKELVDAQLHQDMPFQEVTK 2358
2190 -----PNVQRIFFGDVRSOATNEHFLAARATHLTKGN-----ATKDDVR 2228
2359 LLQVNDNPSRHPVQVNF--NFSRANGEDHARSEDGSLAFNQYRPVQVDSVAKEDLN 2416
2229 QKMALEDMEELLVEPAFTSL--KDRFPGLVEHVEILPKNM-----EAVNE---LSAY 2278
2417 ATVTEL---ESGLRVNPNYATSLFNKSTIQGFHLTYEYLLRQLSELSEAGINEDTQLSLV 2473
2279 RYA--AVVHV-----RGLSDELVLVPEKDDWIDF----- 2306
2474 RPTENGDLHLPLAQSPATTAEQKVASLNQAFERAEFLAAEKIAYVQGDRLASYADLNG 2533
2307 QANQL-----NOKSLGD-----LLKSSDA--AIMAVSK-----IPFEIT----- 2338
2534 QANOLARYIQSVSCIGADGIALMLEKSIDTICILAIWAKAGAYVPLDTPYPPGRVQLI 2593
2339 --AFEROVVASLNSIDENWO-----LSTRSAEGDSSLSVPDIFRIAGEAGFRV 2386
2594 LEEIKAKVLVHSHASKCRHGAKVIAVDSPAETAVSQSOAADLPTIASLGNLA--YI 2651

2387 EVSSAROWSQNGAL-----DAVEFH-----C 2407
2652 IFTSGTSKPGKVLVEQAKVLLRDALRERYFGDCTKHGVLFLNSNVFDESVEQLVLS 2711
2408 CSQERTLVNFTD-----HILRS-DLLTNRPQRLQNRRIAI----- 2444
2712 VLSGHKLIVPAEFAFVADDEFYRMASHTGLSYLSGTPSLQKIDLARLDHLQVVTAAEEL 2771
2445 -----EVRER-----LRSLLP---SYMIPSN 2463
2772 HATYEMKRRFNGPIYNAYGVTTVYVNIITAEFTNISFENALREVLPGRVAVVLTAA-- 2829
2464 VVLDKMLNANGKDRKLSRAKVVPKQOATAALPT-----PISREVEVILCEATEV 2517
2830 -ALQPVFPAVG-----ELYLADGTVTRGYLNQPLLTQDRTIPNPFCKEEDIAMGRFARL 2883
2518 F-----GMKVDITD----- 2526
2884 YKTGDLVRSFRNQOQPQLEYLGRGDLQIKRMGYRIEISEVQNVLTSSPGVREGAVVAKY 2943
2527 ---HFFNLGHSLLA-----TKLSRID--ORLVRITVKDFDH-----PVFADLA 2568
2944 ENNDYTSRTAHSLSVGYTTDNETVSEADILTFMKARLEPTYMPSHLCCLEGALPVTINGK 3003
2569 SVIROGLGLOQVSDGQGDQSAHMAPRTETEAICDEFKVLGFQ-VGITDNFFDLGGH 2627
3004 LDVRR---LPEIINDSA---QSSYSPPRNIEAKMCRWLWESALGMERCGIDDDLFKLGGD 3057
2628 SLMATKLAVRIGHRLDITVSVKQVDFHPVLEFQALALDNLVQSKTNEIVGGRMAEYSPF 2687
3058 SITSLHLVAQIHNOVGCKITVRDIFEH-----RTARALHDHVFMKDS-----DRSNVTQF 3107
2688 QLLFTEPEEFMASEIKPQLELQIIDIYPSQOKAFLEFDHTTARPRPFVFPVDFPPS 2747
3108 R---TEQGPVIGEAPLLP-----IQDWFLSKALQHPMYNHT-----FVY---R 3145
2748 TSPDAAGLTKACESLVNHLIDIFR-TVFAEASGELYQVLSCLDLPQIVTET-NINTA 2805
3146 TPELDVDSLSAAVNDLQOYHDFMRMLKREEVGVSQFAEDFSPAQLRVLVNKVDVGSAA 3205
2806 TNEFLD---EFAKEPVRLGHPILRFTTIKOTKSMRVIMRISHALYDGLSLEHVVRKLMH 2861
3206 VNEILLDQWQSGFNLENGPIS--LCYLGHYEDRSARVWFVHHMAIDTVSMQILVRLDQT 3263
2862 LYNGRSLLP---HQFSRYMQ-YTADGRESHGFRWDIQTONTMTILSDDTVVDGNDAT 2916
3264 LYRNGSLGSGSSFRQWAEALQNYKASDSERNH--WNKLYMETASSISALPT-----ST 3315
2917 CKALHLSKTYNIPSOVLGRSSNIITQATVFNACALVLSRESDSKDV-VFGRIVSGROGL 2975
3316 GSRVRLSRSLs-PEKT-----ASLTQGGIDRQDVSVDLSLLTS-VGL 3355
2976 PVEQDIVGCTNAVPVRAH 2995
3356 ALQHIAPTGP--SMWTIEGH 3373

RESULT 11
US-08-222-617A-12
Sequence 12, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan P.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam

TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 739 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-510-646B-33

Query Match 5.3%; Score 855; DB 3; Length 739;
Best Local Similarity 31.0%; Pred. No. 4.9e-66;
Matches 229; Conservative 150; Mismatches 314; Indels 46; Gaps 17;

QY 1068 VGMGLKVSNIYQHPTLAGISAVVKGDPSTYTIIPKSTHEGPEQSYSGRLWFLDOLDVG 1127
DB 5 LGAEATVRLDFAPTVEALAELEAREVPPALRAADPEHVPLSPFAQRRLWFLDLEGP 64

QY 1128 SLWYLIPYAVRMGPVNVNDALRALAALAEORHETLRTTFEDQDGVGVQIVHEKLSSEMKV 1187
DB 65 NSTYNIPLALRLRGELDRPALQALTDLTHRHESLRTVYPSADGRPYQHVLAPEAEPLG 124

QY 1188 IDLCSDLPPEVFNQEQTPNLSSEAGWRATLLRLGEDDHLITVMHHISDGSIDV 1247
DB 125 VVYPADEAGLAEMAAARHEFDVTSEPLRLVSFLTAPDEHVLILLHLLHAGDGWSLAP 184

QY 1248 LRRDLNOLYSAALKSKDPLSALTPLPIQVSDPAKWKQDFTEQ-----KOLNWKK 1300
DB 185 LTRDLTRATAR-RDGAAP--DWEPLPVQVADYTLWQEMLGSPDDSLGARQLDHWAR 241

QY 1301 QLKDS--SPAKIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFNCNEHTTFSFVLLAAFR 1359
DB 242 SLAGAPEQLELPTDHNRPAAAGHHGRTVPFHEPELHERLSALARSCLASLEFVLAHAAFA 301

QY 1360 AAHYRLTAVEDAVIGTPIANRNPPELEDIGCFVNTQCHRNINDHDDTGTGLINQVATT 1419
DB 302 ALLTKHAGTDPIGSPAGRTDEALDDLVGFVNTLVLRDTSGDPTFRELVARTRAD 361

QY 1420 TAAFENEDIPFERVVSALOPGSRDLSSSTPLAQLIFAVHSOKDLGRFKFQGLSVPPVPSK- 1478
DB 362 LAAYAHQDLPPFEKLVTLPN-QKSLARNPLFQVLLAFQSM-PTAQPVLPGLGVVHPEFVR 419

QY 1479 AYTRFDMFHLFQETD-----SLKGSVNFADFLFKMETVENVVVRVFEILRNLQSSRTP 1533
DB 420 GRAKFDLALAVAEERHADGRSLRGWDEFSTDLFEQATVEALGARLTALLASVAADPDQP 479

QY 1534 VSILPLTDGIVLEKLDVLNVKHVDYPR---ESSLADVTQTQVSAYPDSLVAVDSSCLRT 1590
DB 480 IGRVGILD--PAERHRIILHTWN-DTSRFGADATWPELFQARAAEHDAVALVQEGTETG 535

QY 1591 YTELDROSILAGWLRRRSMPTLTVAVFAPRSCETIVAFFGVKLANLAYLPDVRSPSA 1650
DB 536 YADLNTRANRLARLLRAQIGPEQVVALSPRSADLIVSLAVLKTGAAYLPVDPYPAPAE 595

QY 1651 RVQDILSGLSGPTIVLIGHDTA---PPDIEVTNVFEVRINDALNSNADGFEVIEHDSK 1707
DB 596 RIAYLLQD-GAPALVLTHTSVAAGLPGVQLLVQDGLDD-----VPHDLDLTAERTT 648

QY 1708 P-SATSLAYLYTSGTGRPKGVMIEHRVIRTVTSGCIP-----NYPSETRMAHMATIA 1761
DB 649 PLHPLPAYIVITSGTGLPKGVPPH---RSVASVLVPLIEEFGGLGPGSRVLQFASIS 704

QY 1762 FDGASVEIYSALLFGRTL 1780
DB 705 FDAALWEITLALLSGATLV 723

RESULT 13
US-09-335-409-3
; Sequence 3, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOPHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1410
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-3

Query Match 4.9%; Score 782.5; DB 3; Length 1410;
Best Local Similarity 23.4%; Pred. No. 4.6e-59;
Matches 374; Conservative 247; Mismatches 597; Indels 379; Gaps 64;

QY 1109 VEQSYSGRLWFLDQDLSLWYLIP---YAVRMGPVNVDA--LRRALAALAEORHETLR 1163
DB 77 IQESYMWLR-----TGA--FTVPSGHAYREYDCTDLDVPLRFAFKVVARHDMRL 126

QY 1164 T-TFEDQDGVGVQIVHEKLSSEMKVIDLCGSDLPPEV-----LNOEQTTPFN 1210
DB 127 AHTLPDM---MOVIEPKVDADIEIDLRLGDRSTREARLVSLRDAMSHRIYDTERPPL- 181

QY 1211 LSSEAGWRATLLRLGEDDHLITVMHHISDGSIDVLRRLDNOLYSAALKD-----SKDP 1366
DB 182 -----YHVAVRLDERQTRLVL-----SIDLNNVDLGLS-SIIFKDWLSFVEDP 224

QY 1267 LSALTPLPTQYSDF--AKWKQDFIEQKOLNWKKOLKDSSEPAKIPDFAFALLSGDA 1324
DB 225 ETSPLVLELSYRDYVLALESKKSEAHQSRMSDYWKRIAEPLPPPTLPMKADPSTLK-EI 283

QY 1325 GCYHVT--IDGELYQSLRAFNCNEHTTFSFVLLAAFAAHAYRLTAVEDAVIGTPIANR-- 1380
DB 284 RFRHTEQWLPSDSWGRKLRVGERGLTPTGVILAAAFSEVIGRWSASPRFTLNTLPLNRLP 343

```
QY 1381 NRPELEIIGCVNTQOCHRIINIDHDTFTGLINOVKATTTAAAFENEDIPFERY-----V 1434
Db 344 VHPRVNDITGDTSMVLIDITTRDKSFQRAKRIQEQOLWEAMDCDVSGIEVOREARV 403
QY 1435 SALOPGS--RDLSTPLAQLIFAVHSQKDLGRFKFQGLSEVPVSKAYT-RFDMFHLFO 1491
Db 404 LGIORGALFPVVLISALNQVGVTSIORLG-----TPVYTSOTQOLLDDHOLYE 454
QY 1492 ETDLSKGSVNADELFKMETVENVVRVFEILR-----NGLOSSRTPVSIPLDTGI 1543
Db 455 HDGLVLAVNDVGVFPDLDLDDLEAYVFLRLTEPMEQVRCVSLPPAQLEARASAN 514
QY 1544 VTEKLDLVNKHVDYIPRESSLADVQTOVSAYPDSLAVDVSSCLRIYITELDRSDILAG 1603
Db 515 ATNALLS-----EHTLGLFAARVEQLPMQJLAUVSARKTLTYEELSRSRRLGA 563
QY 1604 WLRRRMPAETLVAFPRCETTVAFGLVKANLAYLPLDVRSPSARVODILSG----- 1658
Db 564 RLREQARPNTLVAVWKEGWQVAVLAVLESAAVYPIDADLPAERIHVLLDHGEVKL 623
QY 1659 -LSGPTVLIGHTADPDIETVNEFVRIRDALNSNADGPEVIEHSTKP-----SAT 1711
Db 624 VLTQFW--LDGKLSWPEGIQ-----RLVSEAGVEG-----DGDQPPMPTQPS 666
QY 1712 SLAVLYTSGTGRPKGMWIEHVLIRTVTS-----GCIPNYPSETRMAHMAITADG 1764
Db 667 DLAVIYVTSGLPKGMWIDHRAVNTILDINERFEIG-----PGD-RVLAISLSFDL 720
QY 1765 ASYEIYALLFRTLVCVDYMTTLDARALKDVFREHVNAAASHVTSQQDPLRV----- 1819
Db 721 SVYDVFGLAAGGTIVPVDASKLRDPAHWAELIEREKVTVWNSVPAL-----MRMLVEHF 775
QY 1820 ---PRRLSRTLMFFLVDTSDAPAL--DAQGLYQGVQYNGYGPTEGWMSTIYDIDS 1874
Db 776 EGRDLSLARSURLSL--SGDWIPVGLPGLQAIKRGVSVISLGCATEASIWSTGYVRN 833
QY 1875 TESFINGPIGRALNNSGAYVVDPEQOLGVGMGLVVTGDLARGY--SDKALDENRE 1932
Db 834 VLSWASIPGRPLRNQTFHVDLEALPRPWPVPGQLYIGVGLALGYWRDEETKRSFL 893
QY 1933 VHTVNDOTKAYTGRVRYRIGDGLIEFFGRMDTQFKRGNRIESAIEAALLRDSV 1992
Db 894 VHPETGE---RLYKTGDLGRY-LPDGNIEMFGRDQIKLGRYVELGETEETLKSPNV 949
QY 1993 RDAVVLQONEDQAPETILGFVWADHDHSENDKGSANQVEGWQHFESGMYSDIGEDPS 2052
Db 950 RDAVIVPVGNDAAKLLLAYV-----PEGTRRAAEQ-----DASLKTE--RIDAR 994
QY 2053 TIGSDFKGTSMYDGSIDFDEMHEWLGTEITRTLHDNRSLGNVLEIGTSGMILFNLSR 2112
Db 995 AHAAEADG---LSDGERVQFKLARHGL-----RRDL-----DGKPVVDLTGQ 1033
QY 2113 LESYVGLG---PSRAAFAFNKATESI-----PSLAGKAKV 2145
Db 1034 DPREAGLDVYARRSVRTLEAPIPFVEFGFRFLSCLSSVPEPDGATLPKFRYPSAGSYVP 1093
QY 2146 QVGTATDIGOVDLHPLDVLVNSVIOYFPSPSEYLAETADTLHLPLNVQRIFFGDRSQAT 2205
Db 1094 QTYAYVKSGRIG-----VDEGFYVHPFEHRLKLUSD----- 1126
QY 2206 NEHFLAARAHTLCKNATKDDVRQKMAELMEDMBEELLVEPAFTLSKDRFPGLVEHVEIL 2265
Db 1127 -----HGTERGA--HVRQNFVDVDEAFAFNLL-----FVGRIDALESL 1161
QY 2266 PKMWEAVNELSAYRAAVHVRGSLGDELVLPEKDDWIDFQANQLNOKSLGDLKSSDA 2325
Db 1162 -----YGSSSRFECL-----LEAGYMAQ-LLMEQAPSCNI 1190
QY 2326 AIMAVSKIPETITAFERQVVASNSNIDEMQLSTIRSSAEGDSSLSPVDIFRATAGRFR 2385
Db 1191 GVCVPGQFNFE-----QVRPVID-----LRHS-----DVI-VHGMILGGR 1223
QY 2386 VEYSSARQWSONGALDAVFHCCSQGRTLVNFPTDHLHLRSDLLTNRLRQLQRNRIATIE 2445

RESULT 14
US-09-568-102-3
; Sequence 3, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1410
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-3

Query Match 4.9%; Score 782.5; DB 4: Length 1410;
Best Local Similarity 23.4%; Pred. No. 4.6e-59;
Matches 374; Conservative 247; Mismatches 597; Indels 379; Gaps 64;

QY 1109 VEQSYSGRLWFLDQDLVDGSLWYLP---YAVRMGRPVNVDA--LRRALAALEORHETLR 1163
Db 77 IQESYWLGR-----TGA--FTVPSCIHAYREYDCTDLDVPRLSRAFRKVVARHMDLR 126
QY 1164 T-TFEDQDGVGVQIVHKEKLSSEMKVIDLGGSDLDPFV-----LNQBOTTPFN 1210
Db 127 AHTLPDM---MQVIEPKVDADIEIDLGLDRSTREARLVSLRDMASHRIYDTERPPL- 181
QY 1211 LSSPAGWRATLRLGEDDHITLIVMHHIISDGWSIDVLRDLNOLYSALKD---SKDP 1266
Db 182 -----YHVAVRLDERQTRLVL-----SIDLNVDLGSL-SLIIFKDWLSFYEDP 224
QY 1267 LSALTPLPIQYSDP--AKWQKQDFIEQEKQLNYWKQKDKSSPAKIPDTDFARPALLSGDA 1324
Db 225 ETSPLVLELSYRDYVLALESKKSEAHQSRMSDYWKRIAEALPPPTLPMKADPSTLK-EI 283
QY 1325 GCYHVT--IDGELYOSLRACFNEHNTTSFVVLLAAFAAHYRLTAVEDAVIGTPIANR-- 1380
Db 284 RFRHTEQWLPDSMGRUKRRVGERGLTPTGVILAAFASEVIGRWSASPRFTLNLITLFRNL 343
QY 1381 NRPELEIIGCVNTQOCHRIINIDHDTFTGLINOVKATTTAAAFENEDIPFERY-----V 1434
Db 344 VHPRVNDITGDTSMVLIDITTRDKSFQRAKRIQEQOLWEAMDCDVSGIEVOREARV 403
QY 1435 SALOPGS--RDLSTPLAQLIFAVHSQKDLGRFKFQGLSEVPVSKAYT-RFDMFHLFO 1491
Db 404 LGIORGALFPVVLISALNQVGVTSIORLG-----TPVYTSOTQOLLDDHOLYE 454
```


Search completed: May 30, 2003, 12:50:06
Job time : 83 secs

```
QY 1604 WLRSSMPAETLVAVFAPRSETIVAFEGVILKANLAYLPLDVRSPSARVODILSG-----1658
Db 564 RLREQGARPNTLVAVVMEKQWQVAVVAVLESGAAYVPIDADLPAERIIHYLLDHDGVEVL 623
QY 1659 -LSGPTVLIGHDTAPDIEVTNVFVRIRDALNSADGFEVIEHDSKTP-----SAT 1711
Db 624 VLTQPW--LDGKLSWPGIQ-----RLLVSEAGVEG-----DGDQPPMPIOFPS 666
QY 1712 SLAVLYTSGTGPRKGMIEHRIIVTYS-----GCPNYPSETMAHMAITAFDG 1764
Db 667 DLAYVIYTSGLPKGMIDHKGAVNTILDINERFEIG-----PGD-RVLAJSSLSFDL 720
QY 1765 ASYEIYSALLGRTLCVCDYMTTLDARALKDVPFREHVNASHVYSSQDVPLRV-----1819
Db 721 SYDVFGILAGGTIVVPDASKLRDPAHMAELIEREKVTVMNSVPAL-----MRMLVEHF 775
QY 1820 ---PRRLSRTIMEFFLVVTTSTAPDAL--DAQGLYQGVQYNGYGTENGVMSTIYPIDS 1874
Db 776 EGRPDSLARSLSLL--SGDWIPVGLPGELOAIRPGVSVISLGGATEASISGYPVN 833
QY 1875 TESPINGVPIGRALNNSGAYVVDPEQOLVGIGVNGELVVTGDLGARY--SDKALDENRF 1932
Db 834 VLSWASIPYGRPLRNOTFVLDLEALPRPVWPGQLYIGVGIALGYWRDEKTRKSFL 893
QY 1933 VHTVNDQTVKAYRTGDRVYRIEFGDGLIEFEGMDTQFKIRGNRIESAEIEAALLRDSSV 1992
Db 894 VHPETGE---RLYKTGDLGRY-LPDGNIEFMGREDNQIKLRGYKVEIGEIEETLKSHPNV 949
QY 1993 RDAAVLQONEDQAPETILGFVVDHDSENDKQSANQVEGWODHFESGMYSOIGEIDPS 2052
Db 950 RDVIVPVGNDANKLLAYV-----PEGTRRAAEQ-----DASLKTE--RIDAR 994
QY 2053 TIGSDFKWTSMYDGSOIDFDEHWEHIGETRTLHDNRSLGNVLEICTGSGMILFNLDNR 2112
Db 995 AHAEEADG---LSDGERVQFKLARHGL-----RRDL-----DGKPVVDLTGQ 1033
QY 2113 LESVGLG---PSRSAAAFVNKATESI-----PSLAGKAKV 2145
Db 1034 DPREAGUDVYARRRSVTELEAPIFVEFGREFLSLSEVPDGTATLPKFRYPAGSTYPV 1093
QY 2146 QVGTATDIGQVDDLHPDLVVLNSIQYFSPSEYLAIEADTLIHLPNVQRIFFGDVRSQAT 2205
Db 1094 QTVAYVKSGRIEG-----VDEGFYVHPFEHRLKLSD-----1126
QY 2206 NEHFLAARATHLTGKNATKDDVQKMAELEDMEELLVEPAFTSLKDRPGLVEHVEIL 2265
Db 1127 -----HGIERGA---HVRQNFDFDEAAFNLL-----EVGRIDATESL 1161
QY 2266 PKMNEAVNELSAYRYAAVHVHRSGLDELVLPEVKDDWIDFOANQLNOKSIGDLKSSDA 2325
Db 1162 -----YSSSREFCL-----LEAGYMAQ-LLMEQAPSCNI 1190
QY 2326 AIMAVSKIPETAFERQVVASLNSNIDEWQLSTIRSSAEGDSSLSVPDIFRIAGEAGFR 2385
Db 1191 GVCVPVGQFNEF-----QVRPVL-----LRHS-----DYY-VHGMGLGR 1223
QY 2386 VEYSSARQWSONGALDAVFHCCSQGRTLYNFTDHLHRCSDLLTNRPLOLONRRYAE 2445
Db 1224 VD---PRQF-----QVCTLGOD--SSPRRATRGA-----PPGREQH--FADM 1259
QY 2446 VRERLSLPSYMPISNIVVLDKMLNANGKVDKELSRRAKVVVKOOTAAPL---PTFP 2502
Db 1260 LRDFLRTKLEYAVPTVFVELDALPLTSNGKVDKALRER-----KDTSSPRHSGHTAP 1313
QY 2503 ISEVEVILCEATEVEFGMK-VDITDHFNFILGGHSLATKLISRIDQRKVRITVKQVDFH 2561
Db 1314 RDALEEILVAVREVILGLEVVGLQSFVDLGATSIHIVMRSLLOKRLDREIAITELFOY 1373
QY 2562 PVFADLASVIRQGLIGLOQPVSQGGQDRSAHMAPRTE 2598
Db 1374 PNLGLASGLRR-----DSRDLQDRPNMQDRVE 1401
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:48:42 ; Search time 85 seconds
(without alignments)
3726.223 Million cell updates/sec

Title: US-09-482-788-2
Perfect score: 16128
Sequence: 1 MEXLTAVDGRQLPPTPSAF.....RVEHLLVEVSKTEGLNSSL 3129

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 10123694 residues
Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCRT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCRTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|----|---------------------|
| 1 | 1713.5 | 10.6 | 2448 | 10 | US-09-815-242-11870 |
| 2 | 1677 | 10.4 | 4342 | 10 | US-09-815-242-5107 |
| 3 | 1632 | 10.1 | 4999 | 9 | US-09-976-059-15 |
| 4 | 1406 | 8.7 | 2397 | 10 | US-09-815-242-12265 |
| 5 | 1375.5 | 8.5 | 4999 | 9 | US-09-976-059-14 |
| 6 | 1374 | 8.5 | 1668 | 10 | US-09-815-242-5654 |
| 7 | 782.5 | 4.9 | 1410 | 9 | US-10-014-717-3 |
| 8 | 623 | 3.9 | 1051 | 9 | US-09-976-059-13 |
| 9 | 612.5 | 3.8 | 1446 | 9 | US-10-166-087-44 |
| 10 | 559 | 3.5 | 1293 | 10 | US-09-815-242-10079 |
| 11 | 538.5 | 3.3 | 1294 | 10 | US-09-815-242-13724 |
| 12 | 475.5 | 2.9 | 891 | 9 | US-09-976-059-18 |
| 13 | 463 | 2.9 | 458 | 9 | US-10-008-016-4 |
| 14 | 434.5 | 2.7 | 1391 | 9 | US-09-994-595-8 |
| 15 | 424 | 2.6 | 1295 | 9 | US-09-738-626-6480 |
| 16 | 415.5 | 2.6 | 401 | 10 | US-09-924-256A-88 |
| 17 | 390 | 2.4 | 485 | 10 | US-09-815-242-5301 |
| 18 | 390 | 2.4 | 485 | 10 | US-09-815-242-12515 |
| 19 | 379.5 | 2.4 | 399 | 10 | US-09-924-256A-86 |

| | | | | | |
|----|-------|-----|------|----|---------------------|
| 20 | 379.5 | 2.4 | 506 | 10 | US-09-815-242-10589 |
| 21 | 377 | 2.3 | 390 | 10 | US-09-924-256A-82 |
| 22 | 376 | 2.3 | 392 | 10 | US-09-924-256A-24 |
| 23 | 369 | 2.3 | 388 | 10 | US-09-924-256A-94 |
| 24 | 369 | 2.3 | 471 | 10 | US-09-815-242-11698 |
| 25 | 366.5 | 2.3 | 390 | 10 | US-09-924-256A-18 |
| 26 | 363.5 | 2.3 | 600 | 9 | US-10-166-087-42 |
| 27 | 352.5 | 2.2 | 396 | 10 | US-09-924-256A-84 |
| 28 | 347.5 | 2.2 | 392 | 10 | US-09-924-256A-92 |
| 29 | 346.5 | 2.1 | 516 | 10 | US-09-815-242-13644 |
| 30 | 334.5 | 2.1 | 512 | 9 | US-10-169-048-56 |
| 31 | 311 | 1.9 | 568 | 9 | US-09-738-626-3950 |
| 32 | 284.5 | 1.8 | 396 | 10 | US-09-924-256A-90 |
| 33 | 277 | 1.7 | 512 | 9 | US-09-738-626-3817 |
| 34 | 256 | 1.6 | 535 | 10 | US-09-947-027-10 |
| 35 | 256 | 1.6 | 535 | 12 | US-10-091-009-10 |
| 36 | 237 | 1.5 | 722 | 9 | US-09-906-419-19 |
| 37 | 237 | 1.5 | 722 | 9 | US-10-119-136-19 |
| 38 | 236.5 | 1.5 | 491 | 9 | US-10-149-485-3 |
| 39 | 233.5 | 1.4 | 550 | 9 | US-09-838-469-31 |
| 40 | 232 | 1.4 | 544 | 9 | US-09-838-469-24 |
| 41 | 231 | 1.4 | 546 | 9 | US-09-838-469-15 |
| 42 | 231 | 1.4 | 546 | 9 | US-09-838-469-23 |
| 43 | 230 | 1.4 | 546 | 9 | US-09-838-469-16 |
| 44 | 229.5 | 1.4 | 3032 | 9 | US-09-836-705-44 |
| 45 | 228 | 1.4 | 547 | 9 | US-09-838-469-25 |

ALIGNMENTS

RESULT 1

US-09-815-242-11870
; Sequence 11870, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11870
; LENGTH: 2448
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11870

Query Match 10.6%; Score 1713.5; DB 10; Length 2448;
Best Local Similarity 28.6%; Pred. No. 2.4e-115;

Matches 586; Conservative 327; Mismatches 750; Indels 385; Gaps* 76;

| | | |
|----|------|---|
| QY | 430 | YDHNVIDSLOTRLLQOFGCHLIK-CLQSPDLSSMAEYVNLMTYEDRAB-IESWNSQPLE-488 |
| Db | 441 | YATDLFDASTYERLACHWRNLLRGIVANP--RQRIGELPLLDAPERRQTLSEWNPAQRECB498 |
| QY | 487 | -VQDFTLIHHEMLKAVSHSPTKTAIOANGDWTYISELDNNVSSKLAVHIKSLGRAOQAIIIP545 |
| Db | 499 | AVQGTPL-CORFEQARQBPQAVALLFDRLSYGELNARANRLAHCLTARGVGAD---VP554 |
| QY | 546 | V--YFEKSWTIASMLAVLKSGNAFTLLDPNDPPARTAOQVVTOTRATVALTSKLHREIVQ603 |
| Db | 555 | VGLALERSLDMVLGAILAILKAGGAYLPDPAAPERLAHILDDSDVRLLLTQGHLLERLP614 |
| QY | 604 | KLVG-RCVVVDELQSVASDDFSSLSKSDLAYVFTSGTGDPKGIMIEH-----RAF658 |
| Db | 615 | ROAGVEVLADGLVDGVAESDPLPTLS-ADNLAYVFTSGTGDPKGPKTLLTHRNALRLF673 |
| QY | 659 | SSCALKFCASLGINSDRALQFGTHAFGACILLEMTWTLINGGCVCII-----PSDDDR710 |
| Db | 674 | SATEAWFGFD---ERDWTL-FHYSADFQSWELFAGLLYGCLVIVPQWVSRSPEDPYR729 |
| QY | 711 | MNSTPSTINRYNVNMMAATPS-----YMGTFSPEDV---PGLATLVLVEGQSSSVNAIW762 |
| Db | 730 | L-----LCRBGVTVLNAQTPSAFQRLMAVACSADMATQOPALRYVIFGGEALDQLSLRPW783 |
| QY | 763 | APKL-----QLLNGYGOSESSICFASNMSTEP-----NNMGRVAGHSWIDP806 |
| Db | 784 | FORCGDQPOLVNMNGITETP-----VHVTVRPVSEADLEGGLVSPIGGTLPIDLSWTLTD838 |
| QY | 807 | NDINRLVPIGAVGELVIESPGIARDYIVPPPEKSPFFTDIPSWYPANTFPDGA--KLRY864 |
| Db | 839 | RDLNP-VPRGAVGELYIGRAGLARYLRRPGLSATRE--VP-----NPFGGAGERLYR889. |
| QY | 865 | TGDLARYASDGSIVCLGRIDSOVKIRQORVELGAIEIHLRQOMPPDDTLIVVEATKRS-QS923 |
| Db | 890 | TGDLARQADGNIBYIGRIDHQVKVRGRIELGEIEAALA-----GLAGVRDVAVVLADHG944 |
| QY | 924 | ANSTSLIAFLGSSYFGNRPESDAHILDHDATKAINIKLEQVLRPHSIPFVICMLELPT983 |
| Db | 945 | VGGTQLVGYVVADS-----AEDAERL---RESLRSLKRHLPDYVYPAHLMLEMRPLT995 |
| QY | 984 | ATCKIDRRRLRMCKDILDKQTOGAIVQQA-PAPIPVADTAAKLHSIVQSLGIDPATV1042 |
| Db | 996 | VNGKLDQAL-----POPDAUSQQARAP---GSELEORIAIAWEILGEV--RV1041 |
| QY | 1043 | NVGATTFELGCGNSITAIKWNMAR-SVGMDLKYSNIYOHPTLAGISAVVAKGDPLSYTLIP1101 |
| Db | 1042 | GLDDNFFELGCHSLLATRVISRYRQEQDASLKAIFERPVLEAFAGQLERTTDVSTIP1101 |
| QY | 1102 | KSTHEGPVEQYSQGRILVDQDLVGSLYLIPIYAVMGRPVNVLDRLRALAALEQRHET1161 |
| Db | 1102 | LADROQPLASFAQERQFWLQWLEPESAAYHIPISALRGLRDVLDALQSRFDSILVARHET1161 |
| QY | 1162 | LRITTF-----EQDGDVGVOIVHEKLESEMKVVIDLCGSDDLPPEVLNQEQTFPNLS1212 |
| Db | 1162 | LRTRFRLEGGRSQVQVPAPVSVSIERQGEGLI-----ERQAIIVVQPFDE1210 |
| QY | 1213 | SEAGWRATLRLGEDDHILITVMHHIISDGWSIDVLRDRDLNOLYSAAKDKSDPLSALTP1272 |
| Db | 1211 | RGPLLRVNLLQLAEDDHVLVLVQGHVIVSGDSQMVVEELVQLY-AAYSQGLDVV--LPA1267 |
| QY | 1273 | LPIQYSDFAKWOKD--QFIEQEKOLNWKQLKDDSSPA-KIPTDFARPALLSDACGVHV1329 |
| Db | 1268 | LPIQYADYALWQSWEAKEKERQLAYWTGLLGGEQPVLELPDRPRPQARSHRGAOLGF1327 |
| QY | 1330 | TIDGELYQSLRACPNBHNTTTSFVLLAAAPRAAHYRLTAVEDAVIGTPIANRNRPPELIEDII1389 |
| Db | 1328 | ELSRLEIVEAVRALAQREGASSEFMLLASFOALLYRYSGQADIRVGVPIANRNREVERLI1387 |
| QY | 1390 | GCFTVNTQCMRINLDHDDTFTGLINQVKATTTAAFENEDIPFERVVSALQPGSRDLSSTPL1449 |
| Db | 1388 | GFVNTQVLKADQGRMGFDELLAQARQALBEAQHODLPFFQOLVLEAQF-ERNASHNPL1446 |

RESULT 2

RESOL 2
US-09-815-242-5107

03 09 613-242-5107
; Sequence 5107, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: OHLSEN, KATH L.
APPLICANT: ZYSKIND, JUDITH W.

APPLICANT: Wall, Daniel

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.

APPLICANT: Carr. Grant J.

APPLICANT: YAMAMOTO, ROBERT T

APPLICANT: YAMAMOTO, ROBERT
APPLICANT: YU, H. HOWARD

APPLICANT: XU, H. HOWARD

TITLE OF INVENTION: Identification of Essential Concepts in

| TITLE OF INVENTION: | IDENTIFICATION |
|---------------------|----------------|
| TITLE OF INVENTION: | PROCESSES OF |

1. TITLE OF INVENTION: PROXY;

FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: U
: CURRENT FILING DATE: 2001-01-01

; CURRENT FILING DATE: 2001-01-01

;; PRIOR APPLICATION NUMBER: 60, 000, 000

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 60,

; PRIOR FILING DATE: 2000-05-2

; PRIOR APPLICATION NUMBER: 60,

; PRIOR FILING DATE: 2000-05-21

; PRIOR APPLICATION NUMBER: 60,

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60,

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60,

; PRIOR FILING DATE: 2000-12-21

; PRIOR APPLICATION NUMBER: 60,

; PRIOR FILING DATE: 2001-02-10

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: Fas

; SEO ID NO 5107

```

, SEQ ID NO 3107
:   LENGTH: 4342

```

```

; LENGTH: 4342
; TYPE: PRT

```

TYPE: PRI
ORGANISM: Pseu

| | | | | |
|-----------------------|-------------------|-------------------|-------------|--------------|
| Query Match | 10.4% | Score 1677; | DB 10; | Length 4342; |
| Best Local Similarity | 23.6% | Pred. No. 3e-112; | | |
| Matches 814; | Conservative 498; | Mismatches 1281; | Indels 852; | Gaps 135; |

| | | | |
|----|------|--|------|
| QY | 65 | LDKQSAIGHAVYDVP-----TQDISRFALAKWEIVNQTPALRAFAFTSDSOKTTSOVL | 118 |
| Db | 690 | IDPOS-----AAYINPGRLRUGELGDAALRASQFLVERHEALRTRFLERDGAALQIDE | 745 |
| QY | 119 | KDSFVFSWMCWSSSSPDEVVRDEAAAA-----SGPCRCNFVLLQDMQTK | 164 |
| Db | 746 | RGFFANQFVDLALAE-----HERAAAAQRREAEAAQOPDELBKGPULLR--VSLVRUDEQ | 798 |
| QY | 165 | KCOLVWTFESHALVDVTFQORVL---SRVFA----- | 191 |
| Db | 799 | EHQLWVTLHHIVADGWSLNLILLDEFSLRYAEACGGQPADLAPLELHYAEFAAQROWLDA | 858 |
| QY | 192 | -----AYKHKEK--DTHRPETPESSDATDTSQSVSVWSMEDNAVSAATHFWQTHLND | 242 |
| Db | 859 | GEGARQJAYWRELRGDT-APVLELATDHPRTAKQASPAARYSLRVDEALARAIREAAL-D | 916 |
| QY | 243 | LNASVFPHL---SDHLMVPNPTTTAEHRITFP--LSQKALSNSAICRTALSILLRYT---- | 295 |
| Db | 917 | HEASVFEMWLLAAQFALLHRSGGGEIRIGVPSANRQRLQTLQGLVGFFINTLVLRGTTPRAR | 976 |
| QY | 296 | -----HSEALFGAVTEQSLPFQKHLYADGT-YQTVAPLVRHVCOSNLRASDVMDAIS | 346 |
| Db | 977 | QPFPAALLGEAREATLGAQANODLPFFQVLAACGGQGLFQVFLFNHQOURL-----SALR | 1030 |
| QY | 347 | SYDRLGHLAPFGLRDIRNTGDNGSAACDPOTVLLVTDGSHVNNGINGFLQQTIESSEHFM | 406 |
| Db | 1031 | RLPCLLADLPWHSRE-----AKFDLQ----- | 1052 |
| QY | 407 | PCNNRALLHCQMESSGALLVAY-YDHNVIDSLQTTTLRLOQFGLHI-----KCLQSP | 457 |
| Db | 1053 | -----LOSEDARGRLTFNFDYAADLDFEASIRRFACQYLLELLRQVAEDPORCL---- | 1101 |

Db 2113 EYTDLDQRLGCLLYSRDLPEPRIARMAGHWQNLLEALLGDPORRIAEPL---FAAE 2169
QY 1548 KLDVNLKVDYDRESSLAD---VFQTVSAYPDSLAVDSSCLTYTTELDRSDILAG 1603
Db 2170 RKOLLACTAG---EAGLDPLHLGFAARVAASQAPALTEFAGOTLSVAELDARSNLR 2226
QY 1604 WLRRESMAETLVAVFAPRSCETIVAFGVKLANLAYLPLDVRSRQVQIL---SGLSG 1661
Db 2227 VLRSHGVGPEVRVGLALERSLEMVGLLAILKAGGAYVLPDPEYLERLQYMIEDSGVR- 2285
QY 1662 PTIVLIGHDTAPDIEVTNVEFVRIRDALNDSNA-----DGFEVIEHDSKPSATS 1712
Db 2286 ---LLLSH-----AALFEALGELPAGVAVRWCLLEEDGPDALDAEDPAPLAALS 2328
QY 1713 ---LAVLYTSGTGPKGVMIEHVR---TVTSGCIPNYPSETRMAMATIAFDGAS 1766
Db 2329 GPQHQAYLIYVTSSTGTPKGVAVSHGEIAMHCAAVIECF-GMRAEDCELHFYSINFDAAS 2387
QY 1767 YEIYSALLFGRTLV-----CYDYMTTLDARALKDVFFREHVNA---ASHVTSSQDVP 1816
Db 2388 ERLIAPLLCGARVYLRAGGOWGABEICEILRAEGVSILGFTPSYGSQAQWLESQGRQLP 2447
QY 1817 LRVPRRLSRTLMFPLVVTDTSTAPDALDAQLYOG---VOCYNGYGPTEGVM---STIYP 1871
Db 2448 VR-----MCITGGEALTGEHLQRIQAFAPAFAPAFNAYGTETVVMPLACLAP 2494
QY 1872 IDSTESTINGVPIGRALNNGSAYVVDPEOOLVGTVNGELVVTGDLGARGYSK-ALDEN 1930
Db 2495 -ERLEEGAASVPIGSVVGARVAYTLDADLALVPOGATGELYVVGAGLARGVHERPALSAE 2553
QY 1931 REVHITVNDQTVKAYRTGDRVYRIGDGLIEFGMDTOFKIRGNRESAEIEAALLRDS 1990
Db 2554 RFPDPPFAEGGRLYRTGDLVRL-CDNGQVEYVGRIDHQVIRGRIELGEIEARLLEHP 2612
QY 1991 SVRDAAYVLAQONEDQAEIILGFV---VADHDHSENDKGSQANQVEGMDHFESEGMYSIDIG 2047
Db 2613 QVRE-ALVLAIDSPSGKQLAGYASVAEQD---EDAQA----- 2648
QY 2048 EIDPSTIGSDFKGTSMYDGSQIDFDMHEWLGETTTTLHDNRSLGNVLGTCGSMILF 2107
Db 2649 ---LREALKT----- 2655
QY 2108 NLDRLSYVGLPERSAAAFVNKATESIPSLA-GKAKVOVGATDIDIGVDDLDPLDLVL 2166
Db 2656 HLKQOLPDY--WPAHLL-----LLASLPLTANGKLDRLALPAPD-----PAL 2696
QY 2167 NSVIQPPSSEYLAETADTLHLPNVQRIFFGD-----VRSQATNEHF- 2209
Db 2697 NRQAYEAPRVLQQLAGVWREVLNVERVGLGDNFFELGDSILSIQVVSRRARGLHFS 2756
QY 2210 -----LAARAIHTLGNKATKDDYR-----QKMAELEDMEEBELV 2243
Db 2757 PRDLFOHQTVQSLAAVAARHSQASQAEQPGVQGSALTPIQHWFFDLPLARHEWNQALL 2816
QY 2244 EPAFFTSKDRFGLVEHVEILPKME-AVNELSAVRYAA-----VVHVRSLGDELV 2295
Db 2817 QPQRAIDL-----GLLRKSQRLVEQHDALRALPRQVDGEWLAQHRPLRQELL 2865
QY 2296 --LPVEK-DWIDFOANQLNOKSL---GDLKSSDAAIMAVSKIPFEITAFERQVVASL 2348
Db 2866 WHVPVQSDFCAELFAKA--QRSLDLEQGPLLRA-----VLVDGPGAGEORLLAI 2913
QY 2349 NSNIDE---WQL-----STIRSAGDSSLSVPDIFRIAGEAGFRVYSARQWSONGA 2399
Db 2914 HHLVWDGVSWRVLLEDLQOYVROFAEC-----AEPALPAKTSAFRDWA--GR 2958
QY 2400 LDVAFHHCCSQ-----GRTLVPFTDHLHRLGSD-----LLTNRPLO 2435
Db 2959 LQA---YAGESSEELGWQALGGOPVEWPCDRP-QGDNREALAESVSLRDPQTRQ 3014
QY 2436 RLQNRRIA---IEVRELRSLPSYMI-----PSNIVVLD----- 2467

Db 3015 LLOQAPRAYRTQVNDLLLTALTALRVLCRWSGOPSTLVQLEHGREALFDDIDLTLSVGVNFT 3074
QY 2468 -----KMLNANGKVRKELSRRAKVVPK-----QOTAAPLPFPPTISE 2505
Db 3075 SAYPLRUTPAQSPGE-SIKAIKEQLRAVPHKGLGVGLVRLYLADPAVRQMAALPTAPIT- 3132
QY 2506 VEVILCEEATEVFMKVDITDHFNNLGHSLSLATKLISRIDORLKRVTIKVDVDHPVFA 2565
Db 3133 -----FNYLQO-----FDQS-FA 3144
QY 2566 DLASVIRQGLGLOQVSDGQODRSAMHAPRTETEAILCDEFKVLQFQVGTIDNFFDLG 2625
Db 3145 D-----ALFQPLDOPTG-----PIHDSQAPLPNELS---VDQGV-----YG 3177
QY 2626 GHSLMATKLAVRIGHRIDTTSVKVDVDPHVPFLQALAI-----LDNLVQSKTNEIVGSGREM 2681
Db 3178 GELV-----LRWTVS-RERYDARTVNELAQYLAELQALIEHCLEDGAGGL-- 3222
QY 2682 AYSVPFOLLFTEDPEEFMASEI-KPQLELQEI-----TODIYPTSTOMOKAFLDHTHTAPR 2736
Db 3223 -----TPSDFPLAQLSQALDALAVPAGIEEDVYPLTPMOEGLLL-HTLLEPG 3269
QY 2737 PFV-----PFVIDPSTSEPDACGLIKACESLVNHLDIRTVFAEAGE-LYQVVLSCLD 2790
Db 3270 TGIYMQDRYRIDSPLDPERFAA---AQWAVARHEALRASFWVNNAGETMLQVHIHPGR 3325
QY 2791 LPIQVETEDINTATNEFLDEFAKEPVRILGHPLIREFTIIKOTKSMRVIMRISHALYDGL 2850
Db 3326 TRIEFLDWSLPELGDGHEERLQALHKREAG-----FDLLEQPPFHLRLIRLGEARYWFM 3380
QY 2851 SLEHVVKLHMYNG--RSL--RSL-----PPHQSRMYQYTA-----D 2883
Db 3381 MSNH-----HILIDAWCRGLLMDNDFEYISALGESRPAANLPTPP---RYRDIYAWLQOR 3431
QY 2884 GRESGHFWRDVIO--NTPMTILSD-----DTVVGDNDATCKALHLSKIVNIP 2929
Db 3432 DLQSRWNSLSESLRGERPTLVPSDRFLREHAGESGMVGVDRYTRLDADAGARLELA 3491
QY 2930 SQVLGRSSNIITQATVFNAAACALVLSRESKDVVFGRIYVSGRO-GLPVEYQDIDVGPCTN 2988
Db 3492 QRYO-----LTVNFAQAAMWALTLLRRFSGERDVLFGVTVAGRPVGMPEMORTVGLFIN 3544
QY 2989 AVPVRAHIES-----DYNOLLHDIQOYLLSLPHETIGTGFSLKNCNTDWPPEAITNFSCCI 3044
Db 3545 SIFLRVOMPAGQORCTVREWLNRFLERNLREHEHLPLVAIQES-SELPKQGPLFDSLF 3603
QY 3045 TYHNFYHPSESQEQRVEMGL--TKFVNTEMDE-----PLYDLAIAAGEVEPDGAGL 3095
Db 3604 V-----FENAPVEVSVLDRAQSLNASSDSGRTHTNFPLTVVVCYPG-----DDLGL 3648
QY 3096 KVTVIKATQLFGRKRVREHLLLEESK 3120
Db 3649 HLSY--DQRYFEAPTVERLLGERK 3671

RESULT 3

US-09-976-059-15
; Sequence 15, Application US/09976059
; Patent No. US20020164747A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
; FILE REFERENCE: 3019-PCT
; CURRENT APPLICATION NUMBER: US/09/976,059
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 4999
; TYPE: PRT
; ORGANISM: Actinoplanes sp.

US-09-976-059-15

| Query Match | 10.1k | Score | 1632 | DB | 9 | Length | 4999 |
|-------------|------------|---|------|------------|-----|----------|------|
| Best Local | Similarity | 22.5% | | Pred. | No. | 7.3e-109 | |
| Matches | 738 | Conservative | 417 | Mismatches | 123 | Indels | 1000 |
| | | | | | | | |
| QY | 43 | SSRIEAKPCTPPOLDMIDCNAL-DKQSAIGHAYVDVPTDID-----ISRFAIAWKEIV | 95 | | | | |
| DB | 1104 | AANIADVYPIAPLQEGIFTHMMADRDSA---DVYVPTVVEFDSRDLDFGLAALQQVV | 1160 | | | | |
| QY | 96 | NOTPALRAFAFTSDSGKTSVILKDSGFVSWMCWSSSSSPDEVVREAAAAAASGPRCNR | 155 | | | | |
| DB | 1161 | DRI-----DVRTSVV-----WQGLREPQVVRHARLPV-----DEV | 1193 | | | | |
| QY | 156 | VLEEDM---QTKKCOLWTFESHALVDVTFQQRVLSRVFAAYKHEKDHTRPETPESSDAT | 211 | | | | |
| DB | 1194 | VRDDLDPVEQLNALGTAW-----NDLS-EAPLVQAVVAA-----RPGDPQ----- | 1233 | | | | |
| QY | 212 | DTDSQSVSVVSMCEDNAVSAATHFWQTH-----LNDLNASVFPHLSDHLMVNPPTTAE | 265 | | | | |
| DB | 1234 | ---RWLAVLRH-----HLVDQHTALDILLLELAAYLAGRGD---LPEVPFPE | 1277 | | | | |
| QY | 266 | HRITFPLSOKALSALCRTLALSILLSRTHSDEALFGAVTEOSLP----- | 312 | | | | |
| DB | 1278 | F-----VAHTRLGVPREEHRYFAGLLGDVTEAPTGLGLDVHSGGLASA | 1322 | | | | |
| QY | 313 | DKHYLDGTV--QTVAPLRVHCQS-----NLRASDVMDAISYDDRLHGLHAPGLRDIRN | 365 | | | | |
| DB | 1323 | QAHLRLDGLPLGRVAAFAREHGVSPATLFLHAWARVLGTLAGRDDVVGFTVLFGRM---- | 1378 | | | | |
| QY | 366 | TGONGSAACDFQTVLLVTD-----GSHVNNINGIFLQOITTESSHFMPCNNRALLLHCQ | 418 | | | | |
| DB | 1379 | ---NSGAGADRVPLGFTINTLPVRVRLGAPVGDALDGLRQOLIE-----LIAH-- | 1422 | | | | |
| QY | 419 | MESSGALLVAYVDHNVID-----SLOTTRLLQ----- | 445 | | | | |
| DB | 1423 | --EHAFLAVAAQAANLFGRLPFTSIFNYRVARGAEPAGAALDGIILLASBDLNYPLAVA | 1480 | | | | |
| QY | 446 | -----QFGLHITKLCQSPDLISS-----MAEVNLMTEYDRAEI | 477 | | | | |
| DB | 1481 | VDAEGDTFSLTVDAVAPADPVOVGELLVLTALRNLTRAEANPGTPLAAVGVGEDELSRV | 1540 | | | | |
| QY | 478 | ES--WNSOPLEVDQTLIHHEMLKAVSHSPTKTAIOAWDGDWITYSELDNVSSRLVAHVKSLG | 536 | | | | |
| DB | 1541 | VSGWNTARRVROASVPELFAERVAAPGAPAAAGDLRWYADLDARSALARSILVAAG | 1600 | | | | |
| QY | 537 | LRAQALIPYVFEKSKWVIASMLAVLKSNAFTLIDPNDPPARTAQVVTOTRATVALTSK | 596 | | | | |
| DB | 1601 | VTAESPVV--VALERSADVLTAFILAVAKAGGVFPVDLSWSPRVDAVIADCAAMIAVADR | 1659 | | | | |
| QY | 597 | LHRETQVKLGRCWVDDLELQSVASDDFSLLTKSQ---DLAVFTSGTSGDPKPGIMI | 653 | | | | |
| DB | 1660 | -----PMTGLTVVPAAR-----ACDPAVALPPRPLPCAAAYRWYTSSTGRPKGVVT | 1706 | | | | |
| QY | 654 | EHR-----AFESCALKFGASGLGINSOTRALQFQTHAFGACLLLEIMTTLINGGCYCI--PSD | 707 | | | | |
| DB | 1707 | THQNVVLDVTDRC---WGPT-----PRVLFHAPHAFDASSFELWVPLLTGGTVVWAPGE | 1757 | | | | |
| QY | 708 | DDRMNSIPSIINRYNVNMWMTATPSYMGTFSS--PEDVPGLATLVLVGEQM--SSSVNATW-- | 762 | | | | |
| DB | 1758 | SIDTGVRLQRIARHELTHVHTVAGTLRLVAEDPSCFAGLFEVLTGGDVWPAEAVRRVLDA | 1817 | | | | |
| QY | 763 | APKLLOLLNGYQSESSSICFASNMSTEPN--NMGRAV-GAHSWVIDPDNDINRLVPIGAV | 818 | | | | |
| DB | 1818 | NPGRVVRQLYGPTE-VTLCATQHVVRPEVPLPIGRPLDNRVTVLDD--GLLOQVPVPGVT | 1874 | | | | |
| QY | 819 | GELVIESPGIARDYIVPPPPPEKSPFFTFDIPSWPANTFPDGAKLYRTGDLARVASDGSIV | 878 | | | | |
| DB | 1875 | GELYIAGAVGARYADMPGTTAERFVAD-----PFTAGGRLYRTGDLVRWTEGELY | 1926 | | | | |
| QY | 879 | CLGRIDSQVKIRGORVELGAETHL-----ROOMPDLLTIVVEATKRROSAN | 925 | | | | |
| DB | 1927 | FAGRADQVKIRGYRVEPGVEAVLAALPGVSOAAVIVREDVPDQ-----KKR----- | 1973 | | | | |

| | | | | | | |
|----|------|-------------------|------------------------|------------------|---------------|------|
| Qy | 926 | STSLIAFLGSSYFGNRP | SDAHLDHDAIKAINIKLEQVLP | PRHSISFYSICMLEL | PRPAT | 905 |
| Db | 1974 | --LVAYLVAAP----- | ETVEAARAHAEORLP | SVLIVPSAFVQLD | ALPTGN | 2016 |
| Qy | 986 | GKIDRRRLRMGKDILD | KOTQGAIVQOAPAPIPV | FADTAAKLHSIWQSLG | -----IDP | 1039 |
| Db | 2017 | OKVDRAAL----- | PAPLGFEGAGRAPADAR | EELVGAAFAEVL | LDL | 2056 |
| Qy | 1040 | ATNVGATFFELGNSIT | AIKMYNMARSVMGLKVS | NIYQHPTLAGISVVK | GDPLSYTL | 1099 |
| Db | 2057 | GRVCPDDFFALGGHSL | ALLAVERLRQGLGVS | RAVDFARTPAALT | TRGDCGADRP | 2116 |
| Qy | 1100 | IPKSTHEGPVEQSYSG | RWFLDQLDVGSUWYLI | PYAVRMGPVNV | DALRRALAE | 1159 |
| Db | 2117 | LRAGARPARLPLSYA | QORRLWFLAQLEG | SATYNIPVALR | LEGDLDRD | 2176 |
| Qy | 1160 | ETLRTTFEDODGVCV | QIVHEKLESE--MK | VIDLCGSDLP | FEVNLNQBOT | 1217 |
| Db | 2177 | EVJLRTVFTADGEP | WCHILDPARAEPALP | VVDVPAGRVE-- | EAAVEAAAYAF | 2234 |
| Qy | 1218 | RATLLRLGEDDHIL | TIIVMMHIIISDGS | IDLVRDLNOLY | SAALKDKSDPL | 1277 |
| Db | 2235 | RAVLLAPGDGTHVL | VLVLVHHIADGWS | MRPLARDLATAY | --ARRRGAPES-- | 2291 |
| Qy | 1278 | SDFAKWOKQOFIEQ | -----ERQNLN | YWKQKLDSSPAK-- | IPTDFARPALL | 1328 |
| Db | 2292 | ADYALMQRDLLGSD | SDPASLISRQIAH | WRERL-DGVPE | ELDLPADRPRA | 2350 |
| Qy | 1329 | VTIDGELYQSLRAF | CNEHNTTSFVLLA | AFRAHYRLTAV | EDAVIGITPIAN | 1388 |
| Db | 2351 | AETPADVHRSLR | RAADHGATVMTLO | QAAYAVLLSR | LGAGTGDPVIG | 2410 |
| Qy | 1389 | IGCFVNTQCMRIN | DHDTFTGLINOK | YKATTTAAFEN | EDIPFERVVSAL | 1448 |
| Db | 2411 | VGFENVTLVRLT | DGTDPRLTDVGL | QVREILTRALAH | QDVPFEKLVEEL | 2469 |
| Qy | 1449 | LAOLIFAVHSQKDL | GRFKFQGLSESV | VPYSKAY--TRF | DME----- | 1501 |
| Db | 2470 | LFQVMVTLDCGG | PDGA--ELPGL | AMSVPVTGA | VPKAPKPDLDL | 2528 |
| Qy | 1502 | FADBLFKMETVEN | VVVVFEILRNG | LOSSRTPVSIL | PLTGTGIVTLE | 1561 |
| Db | 2529 | AAADLFDAGTAAR | LAGVLSRVL-- | GVLLA--DPR | RLAEVDPLEA | 2585 |
| Qy | 1562 | --ESSLADVFQTV | SAYPDSILAVD | SSCLTYLTDL | RQSDTLACGL | 1619 |
| Db | 2586 | LPEITAAALVABO | CARTPGAVVGP | PDASITYAEL | DERAARIKWL | 2645 |
| Qy | 1620 | APRSCETIVAF | FGVLKANLAYL | PLDVRSPAR | VDQILSLGSGPT | 1679 |
| Db | 2646 | MERSAELVALLG | VMRAGAAVVP | DPAYPAPER | IRFVVTDARA | 2702 |
| Qy | 1680 | NVEFVRTDALN | SDNADGFEVIE | -----HDS | TKPSATSLAY | 1734 |
| Db | 2703 | GVPCLAIT---- | DDPAAAAAP | EAEPEDGDP | GGAAGRPDP | 2757 |
| Qy | 1735 | VIJRTVT--SGC | IPNYPSETRMA | HMATIAFDG | ASYEISALL | 1793 |
| Db | 2758 | NVALLATRLPL | FGAGDEVMS | FWHSAFDF | SVWELMGAL | 2817 |
| Qy | 1794 | KDVFFRCHVNA | ASHVTS | SSODVPLRV--- | PRLSRTL | 1850 |
| Db | 2818 | WELVVRGVT | VLQTPSAFAQL | MAAGDDDRD | ALRFVVF----- | 2867 |
| Qy | 1851 | G-----VQC | YNGYGTENG | VMSTIYPID | STESFINGV | 1902 |
| Db | 2868 | GWLARRDPK | RLPNVMYGIT | ETTHTT | YQHIAPGTT--- | 2924 |
| Qy | 1903 | VGTVMGELV | VTGDGLARGY | SDK--ALDEN | RFVHTVNDQ | 1961 |
| Db | 2925 | VPAGVPEV | YARGQVARGY | IGRPLG | FAERFV-- | 2982 |

QY 1962 FFGRMDTQKIRGNRIESAETEAALLRDSSVDAVVLQONEDQAPETILGFVADHDS 2021
 DB 2983 FAGRSDDIKIRGRIEPEVEAVLAAAGPGVQAIVIRE----- 3032
 QY 2022 NDKGOSANQVGGWQDFHESGMYSDIGEIDPSTIGSDFKGMTSMYDGSQIDFDEHWEHLGE 2081
 DB 3023 -----DVP----- 3026
 QY 2082 TTRTLHDNRSLGNVLEITGSGMILFNLDLSLESYVGLPEPSRAAAAFVNKATESIPSLAG 2141
 DB 3027 -----DKRLVAV----- 3034
 QY 2142 KAKYQVGTATDIGOVDDHLPLVVLNSVIOVFPSSYLAEIADTLHLPNVQRIFFGDVR 2201
 DB 3035 -----VG-----GDA- 3039
 QY 2202 SOATNEHFLAARAIHTLGKNAKDDVRQKMAELEDMEELVEPAFTSLKDRPGLVEH 2261
 DB 3040 ----- 3039
 QY 2262 VEILPKNEAVNELSAYRAAVHVHVSGLGDELVLPEKDDWIDFQANQLNOKSLGDLK 2321
 DB 3040 ----- 3039
 QY 2322 SSDAAIMAVSKIPPEITAFERQVVASLNSNIDEMQLSTIRSSAEGDSSLSPVDIFRIAGE 2381
 DB 3040 -----ETLSHAQ----- 3047
 QY 2382 AGFRVESSARQWSONGALDAVFFHCCSQGRTLVNFTDHLRGSLLTNRLQRLQNR 2441
 DB 3048 -----QR----- 3049
 QY 2442 IAIEVRERLSLPSYMPISNVILDKMPLNANKVDKELSRKAKVVPKQOAAAPLTF 2501
 DB 3050 -----LPGYLVPAFAVELDKRLPTVNGKLDNR-----ALPVDY 3083
 QY 2502 PISEVEVILCEEATEVEGCMKVDITDHFNLGGHSLATKLISRIDQRLKVRITVKQVFDH 2561
 DB 3084 ----- 3083
 QY 2562 PVFADLASVIRQGLGLOQPVSDGQDQRSAHMAPRTEAILCDEFKVLGFO-VGTDN 2620
 DB 3084 -----GRDAGGRAPANAREVLCRAFAEVLGVERGVEDD 3119
 QY 2621 FFDLGGHSLMATKLAVRIGHRLDTTSVKDVPDHPVLFOALALDNLVQSKTNEIVGGR 2680
 DB 3120 FFALGGHSLVLSVERL-RQGISVPVRFALFTTPPAGLA-----EAVG--D 3164
 QY 2681 MAEYSPQLFTEDPEEFMASEIKPOL-----ELOEI-----IQDIYPTQ 2721
 DB 3165 GAVVVPNLI-----PED--AAELTPEMLPLADLTADLAVVASVPGGAANIADVPLAP 3218
 QY 2722 MQKAFLEDHTA-RPRFV---PFYIDPSTSPDAAGLKACESLVNHLIDERT-VFAE 2776
 DB 3219 LQEGIFHHMAADRSADVVTVVEFDSRDLG--GFLAALQVQVDRDVTYRTSVWQ 3276
 QY 2777 ASGELQVVLSCDLPLIOVETEDNTATNEFLDEFKPEVRLGH-PLIRFTIILKQTS 2835
 DB 3277 GLREPQVVRHARLPIDEVELHEGTDPA--EQIALATERVOLDRAPLIRTTAAVPGS 3334
 QY 2836 MR--VIMRISHALYDGLSLEHVVRKHLMLYNGR--SLLPHQPSRYMOMYADG--RESGH 2889
 DB 3335 GRWALLRIHVLVODHTLVDLLGELRAFLFEGRGDELPEVPFPREFVAQARLGVPREHE 3394
 QY 2890 GFREDVION-TPMTILSDTVDVGNDACTKALHLSKIV--NTPSQVLRSSNI-ITQATV 2945
 DB 3395 RYFAELGVDVETAPYGTETEVEHGDGA--AVHSREVDDDLAARHLRLARSIGVSPAAL 3452
 QY 2946 FNAACALVLSRESDSKDVFGRIYVSGRQGLPVEYQDVGCTNAVPVRAHIESSDYNQLL 3005
 DB 3453 FHLARVLCAGSRDDVFGTVLFCRMNSGAADRQVGLFINTLFPVRVFLAAGSTRDAL 3512
 QY 3006 HDIQDY--LLSLPHETIGFSDLKRNCTDWEPAITNFSCCITVHNFEYHPESQEOORVE 3063

DB 3513 TGLRDQLAGLLVHCHAPLALAQRAAGITDGSPLFASI-----FNVRHNDQDPAASAGLE 3566
 QY 3064 MGVLTKFVNIEMDEPLYDLAIAGEVEPDGAGLKVTVIA 3101
 DB 3567 -GIRTVISAETNPL-DASI--DVTGDRFAITVNAVA 3600

RESULT 4

US-09-815-242-12265
 ; Sequence 12265, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA 011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12265
 ; LENGTH: 2397
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-12265

Query Match 8.7% Score 1406; DB 10; Length 2397;
 Best Local Similarity 24.2%; Pred. No. 6.8e-93;
 Matches 571; Conservative 394; Mismatches 903; Indels 495; Gaps 94;
 QY 212 DTSQSVSVVSMSCEDNAVSAHFQWTHLNDLNASVPPHLSHLMVNPPTTTAEHRTTFP 271
 DB 176 DKNQDASHI-----ALDSNVRLEKNSDI-----HDSYLPKHPPEQALYQ-TYL 221
 QY 272 LSKALSNSAICTALISILSRVTHSD--EALFQAVTEQSLPDKHYLDGTYQTVAPLR 329
 DB 222 IDD--MTSIDMASLAVSVYLANHIMSQQHVDVTLGIHVPSHLPNDLH-----GNIVPLT 272
 QY 330 VHCQSNLRASDVMDATSSYDDRLGLHAPFLGRDIRNTGNGSACDQFQVLL--VTDGSH 387
 DB 273 L-----TIDAKVQCRFTT-----DFNKCVLQNMNSLOQC 301
 QY 388 VNNGINGFLQQTITESH-FMPCNNRAL-----LLHCQMESSGALLVAYYDHNV----- 434
 DB 302 AKSSLS--LETIFHCYHQMPCNDVEDVHQIHAHTSLADTEIFPHQHGFKIYNAA 359
 QY 435 ---IDSLOTRRLQOQGHILKQSLDLSMAEVNLMTEYDRAEIESNSQPLEVOD--- 489
 DB 360 YDLSLSTIETSLVRNLYLQITEENGKRTVTDELNLMTERD---IQLYDDINLSLPDLD 416
 QY 490 ----TLIHHEMLKAVSHSPTKTAIQWGDW--TSELDNVSSRLAVHIKSLGRLAQQA 543

Db 417 AOTVTLFEQO---VEATPNHVAQ-FDGVFIYQTLNARANDLAHR-----LRNOYGV 466
 Qy 544 IP-----YFEKSKWVIAVLKSGNAFTLIDNDPPARTAQVVTOTRATVALTSKLH 598
 Db 467 EPNDRAVIAEKSEIMIIAMIGVLKAGAYVPIDPNYPSDROEYILKDATPKVVIYQAL 526
 Qy 599 RETVOKLVRCVVDDQLQSVAS--DDFSLTKSQDLAYVIFTSGSTGDPKGMIEHR 656
 Db 527 YENSKONINH---ID---LNKIAWKINIDNSECNTLEDHAYVIYTSGTGPNKGLTIPHR 580
 Qy 657 AFSSCALFEGASLGINSOTRALQFTHAFGACLLIMTTLINGGCVCIPSDDDRNW--SI 714
 Db 581 GIVRLVHO--NHVYPLNEETILLSTGIIAFDAATFEIYGALLNGGLIIVAKKEQLLNPIAL 639
 Qy 715 PSFINRYNNMMATPSYMGTFSPEDVPGLATL--VLVGEOMSSVNAIWAPEKL----- 766
 Db 640 EOLINENDVNTMWTLSLFENQIASERIEVPLKYLILGGEV---LNKAWDILLNORPKH 696
 Qy 767 -OLLNGYGOSESSICFASNKSTE--PNN--MGRV--GAHSWVIDPNDINRLVPICAGVEL 821
 Db 697 PQIINGYPTENTTTTNYINIPKNVPIPIGKPIGLTHVIMQG---ERRCGVGIPGEL 753
 Qy 822 VIESPIARDXIVPPPEKSPFTDIPSWYPANTFPDGAKLYRTGLARYASDGSIVCLG 881
 Db 754 CTSGFLAAGYLNQPELADPKID-----SNI---NOLMYRSGDIVRLLPDGNIDYLY 804
 Qy 882 RIDSOVKIRGORVELGALETHLRQOMDDLTIVVEATRKSQANSTSLIA-----FLI 934
 Db 805 RKDAQVKIRGFIEELSEV-----HALERIQINKAVVIVQNHDDQOYIV 849
 Qy 935 GSSYFNGNRPDSAHILDHDATAINKLEQVLPKPSIPSYICMLPRTATGKIDRRRLR 994
 Db 850 --AY--EAMHTLSHNKIKS---QLRMTLPEYMPVNFHIEQIPITINGKLDKALP 900
 Qy 995 IMGKDILDQOTGAIVQAPAPIPVFADTAALKHSIWWQSLCIDPATVNVGATFPELGN 1054
 Db 901 IM--DYVD-----TDAYVAPSTDEHLLCQIFADILHN--QVGIHNFPELGGH 946
 Qy 1055 SITALKMNV--MARSYGMOLKYSNIYQHPTLAGIS--AVVKGDPLSYTLIPKSPHEGPVEOS 1112
 Db 947 SILKATLVNRIEASTGKRLQIGDLDLQKPTVFEQAQIAKVOEQNEVEIPEAIVKDDYLS 1006
 Qy 1113 YSQGRWFLDQDVGSLWLPIDYAVMRGPNVNDALRALAALQORHETLRTTF--EDDQG 1171
 Db 1007 SAKRMYLLWKSNNHRTVYVNPFLWRLSSELNVAOLROAVQLIARHEILRTQYIVVDD 1066
 Qy 1172 VGVQIVHEKLSKEMKVIDLGGSDLDPPFVLNOEQT-----TPNLSSEAGWRATL 1221
 Db 1067 VQRIVADVAD-----FEVNTHTDEQIEIMRQFVAPFNLEKPSQIRVRY 1112
 Qy 1222 LRLGEDDHILTIVMHHIISDGSVIDLRRDLNQLYSAALKSDKPLSALTPLPIQYSDFA 1281
 Db 1113 IRSPLHAY-LFIDTHHIINDGMSIQLMNDNALYQHL-----LLPLKLOYKDY 1162
 Qy 1282 KWQKQDFEIQEKOLNWKQLKDDSP-AKIPDFARPALLSDGACVHVITDGLYQSLR 1340
 Db 1163 EKMSHRDTKHQO---YWLISOFEVPIILSLPTDYVRPNKTTNGAMMSFTMNQOTROLL 1220
 Qy 1341 AFCNHNHTSFVLLAAFAAAHYRLTAVEDAVIGTPIANRNRPEDIEDIGCFVNTQCMRI 1400
 Db 1221 KVEKHQITDFEFMSVNTLLSRYARKDDVVVGVMSARMHKGAEQMLGMEFANTLVYRG 1280
 Qy 1401 NIDHHDFTGTLINQYKATTTAAFENEDIPFERVVSALOPGSRDLSSTPLAQLIFAVHSOK 1460
 Db 1281 OPSDPKMTQFLQEVKENSLEAYEHQEPFECVLNDLQ--SHDASRNPLFDVWMLVQONNE 1339
 Qy 1461 ----DLGRFKOGLESVPVPSKAYTRFDMFHLFOETSLKGSVNFADFLKMETVENVV 1516
 Db 1340 TNHAHFGHSKTHIQ----PKSVTAKFDLSFIIEEDRDDYITINIEYNTDLXHSYVRHG 1395
 Qy 1517 RVFFELRNGLOSSRTPVSILPLTDGIVTLEKLDLVN----KHVDYPRESSLADVFQTO 1572

Db 1396 NOCM-IMIDYILKHQDHTLQICDPNG--TEELNNWNTHVNDMLNPNVGNKSIISYFNEV 1452
 Qy 1573 V5AYPDSLAVDVSSCRLTVTELDROSDILAGLRRRSMPEAETLVAVFAPRSCETIVAPFG 1632
 Db 1453 VSRQGNHVALVNDLTMETLRYVDAIAHLLSNGVNGCORVALFTTERSEMTAAMLA 1512
 Qy 1633 VLKANLAYLPDVRSPSARVODILSGLSGPTIVLIG--HDTAPPOIEVTNVE-FVIR-- 1687
 Db 1513 TVKVGCASYIPIDIDFPKROGAILEDKAVTAVMSYGIETETTLPLVIOLENAKGFVESKEN 1572
 Qy 1688 ---DALNSNADGFEVIEHDSKPSATSLAYVLYTSGSTRGPKGVMEHRVIRIIVTSC 1744
 Db 1573 EYDLDHGDQLENTAMLNDE-----MYAIYTSGTGMPKGVAIQRLNLLVHA-- 1621
 Qy 1745 IPNYSEYRMA-----HMATIAFGASYEISALLFGRTLVCVDVMTTLDARALKDVF 1798
 Db 1622 ---WSTELQDNEVFLQHANIVFDASVMEIYCCLLNGHTLVIPOREERNVPAOLOLIN 1678
 Qy 1799 REHVNAASHVTSSQDVPLVRPRRLSRTLMPFFL--VVTDSAPDALDAQGLYQVQC-- 1854
 Db 1679 KHRVTVAS-----IPLQCMVMEDEFYIEKLITGGATSTASFVK--YIEKHCGT 1724
 Qy 1855 -YNGYGPTEGYMSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEQQLVGVGMCELV 1913
 Db 1725 YFNAYGPSESTVITSWQHCCGLDIPETIPGRPLSNIOVIYIMS-EGLLCGIGMPGELCI 1783
 Qy 1914 TGDGLARGY-----SOKALDENRFVHITVNDQTVKAYRTGDRVRYRIEDGLIEPFGRM 1966
 Db 1784 AGDSLAIQVYNRPELMAOK-WQNPF-----GKGLYHSGDLARY-TSDQIEFLGRI 1834
 Qy 1967 DTQFKIRGNRIESAIEAALLRDSSVRDAAVVLOQNEOQAEILFVVAHDHSENDKQ 2026
 Db 1835 DKQKVGNGYRIEIDETENAILAIRGISDCVVTVSHFDTH--DILNAYVYGEQVQEDLQ 1892
 Qy 2027 SANQVEGMDHPESGMYSDIGEIDPSTIGSDFKGWTSMYDGOIDFDEHWEHLGETRTL 2086
 Db 1893 YLN-----DHLPKYMI-----PKTI-----THIDCMPL-----PTNDK 1920
 Qy 2087 HONRSLGNVLEIGTSGMILNLDLSRLSYVGLPSRAAFVFNKATESIPSLAGKAKVQ 2146
 Db 1921 VDTTRLNPFSPi-----QOSNKVYSEPS-----NEIEQTFVDVFGVQLQ 1960
 Qy 2147 VGTATDQGVDD-----LHPDLV-----LNSVIQY-----PPSSE 2177
 Db 1961 ----NDVGVDDEFELGGSNLEAMLVSDLKRFHHISMOTLYQYKTVQIVNYMHQNOQ 2016
 Qy 2178 YLAEIADTLIHLPNV--ORIFFGCDVRSQATNE-----HFLAARAIHTLGKNATK 2224
 Db 2017 SLVALPDNLSELQINIVWTRYNLGILEDSLSHRPLGNTLGTGATGFLGAYLIEAL----- 2070
 Qy 2225 DDVROKMAELEDMEBELLYEPAFFTSKDRFPGLVEHVEILPKNMEAVNELSAYRYAUV 2284
 Db 2071 QGYSHRIYCFVRADNEEIAWYKMLTNLDYFS--EETVEMLSNIEVI----- 2116
 Qy 2285 HVRSGLG-----DELVLPEKDDWI-----DFOANOLNQSLSGLDKSSD---A 2325
 Db 2117 ----VGDFECDDVVVLPENMDTIHAGARTDHFQDDDEFKYNVQGTVDVIRLAQOHA 2171
 Qy 2326 AIMAYSKI-----PFEITAF--ERQVVASLNSNID-- 2353
 Db 2172 RLIIYSTISGVYFDIDEDVTFSEADYVKGQLLISPYTRSKFYSELKVLAEVNNGLDGR 2231
 Qy 2354 -----EWOLSTIRSS 2363
 Db 2232 IVRVGNLTSPYNGRWHMRNKTN 2254

RESULT 5

US-09-976-059-14
 ; Sequence 14, Application US/09976059
 ; Patent No. US20020164741
 ; GENERAL INFORMATION:
 ; APPLICANT: Farnet, Chris

APPLICANT: Zazopoulos, Emmanuel
APPLICANT: Staffa, Alfredo
TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
FILE REFERENCE: 3019-PCT
CURRENT APPLICATION NUMBER: US/09/976.059
CURRENT FILING DATE: 2001-10-15
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 14
LENGTH: 4999
TYPE: PRT
ORGANISM: Actinoplanes sp.
US-09-976-059-14

Query Match 8.5%; Score 1375.5; DB 9; Length 4999;
Best Local Similarity 28.0%; Pred. No. 4e-90;
Matches 473; Conservative 267; Mismatches 690; Indels 257; Gaps 55;

QY 277 LSNSAICRTALSILLSRYTHSDEALFGAV-----TEQSLPFDKHYLADGYQTVAP 327
DB 3471 VSPATVFLAWARVLSVIAGDDVVFGTILFGRMNSGAAAEVFP-----GLFINTLP 3522
QY 328 LRVHCQSNLRASDYMDAISYYDDRLGHL-----APFGLRDIRNTGDMGSAACDFOTLLV 382
DB 3523 VRV-----RLNCTSVGEALTALRDQMAELMAHEHAPLALAQAGGVPAGS-----PLFTSLFN 3575
QY 383 TDGSHVNGINGFLQQTITESHSFMPCCNNRALLLHCQMESSGALL-----VAYYDHNVID 436
DB 3576 YRNVAGGGDGGGALLEGVTPVLRHTTNNPVVV-----SVDDDDTSFDLVVEAVAPAEAGRVG 3632
QY 437 SLOTRLLQFGLHKLKQSLPDLSSMAEYNLMTEYDRAEIE-----SWNSOPLEVQDTLI 492
DB 3633 RLMECELAELVGLAGAPETPL-----SRVRVI-----DEAEIERVYHWNDDTARPVVSSV 3684
QY 493 HHEMLKAVSHPTKTAQMDGWMTYSELNDVSSRLAVHIKSLGLRAQQAIIPIVFEKSK 552
DB 3685 PALFAEQVAAPADATAVGEGVSWSYRELDARSALRSVAAGVGVEPVV-VALERSP 3743
QY 553 WVTASMLAVKSGNAFTLIDNPDPARTAOVTTQTRATVALTSLKHRETIVOKLGRVCVV 612
DB 3744 EVLSAFLAVAKAGGVFPVLDLSPQARIDAVADCAARAVADR-----PMSGLTWP 3796
QY 613 DDELLQSVASDDFSSITKSDGL-----AYVIFSGSGDPKGMIEHRAFSKALKFGAS 668
DB 3797 ADQV-----GSAVVLPAQVPGAAVRYMTYSGSGRKPQVYTHQNLVDLA----- 3843
QY 669 LGINSDF-----RALQFGTHAFACILLEIMTTLINGGCVCI-PSDDRMNSIPSFNRY 721
DB 3844 ----TDFCWGPTPRVLFPHAFADASSYEIWPVLLNGTGVVAPORSIDATVLRDLRGH 3899
QY 722 NVNWMATPSYMGTFSPEDVPGLATLVLGQWSS-SVNAIWA--PKQLLLNGYGOSSESS 778
DB 3900 ELTHVHVHTAGLLRLVLDPSCFAGLTEVLTGGDAVSAEAVRRVREANPGLRVQLYGPTE-V 3958
QY 779 SICFASNMSTEPNNMGRAV-GAHSWITDPNDINRLVPIGAVGELVIESPGIARDYIVPPP 837
DB 3959 TLCAQHLLVDGVPICRPLDNTRVYVLD--DLQPPVPGVTGELYVAGAGLARGYAGMPG 4016
QY 838 PEKSPFFTDIPSYPANTFPDGAKYRTGDLARYASDGSIVCLGRIDSQVKIRGORVELG 897
DB 4017 LTAERFVAD-----PFSVGGRLYRTGDLVRMTDDGVLHFAGRADDDQVKIRGYRVEPG 4068
QY 898 AIETHLRQOMPD--DLTIVVEATKRSOSANSTSLIAFLIGSSYFGRNRPNSDAHILDHATK 955
DB 4069 EVEAVLAQH--PDYSQAVVVY-----REDTPGDKRLVAVVVGDD----- 4105
QY 956 AINIKLEOVLPRHSIFSYICMLELPTATGKTDRRLRMGRKIDLDKOTQGAIVQOAPA 1015
DB 4106 -VEAYAQERLPGVLSFSAFVHLDALPLTNSQKVRDRAAL-----PA 4144
QY 1016 P-----IPVFADTAALKIH-SIWVQSLGIDPATVNVGATFFELGNGSITAIKMYNMAR 1066
DB 4145 PSVESGVGRAP--ADAREELMCAAFAEVLDD--RVGVDDDFALGSHSLVLYVRGRIR 4200

QY 1067 SV-GMDLKVSNIYQHPTLAGISAVVKGDPGLSYTLIPKSTHEGVPQSYSGRLWFLDQID 1125
DB 4201 QVGEVGSARLVDFDARTPAGVVVRLSEGTAAREAVRARVPRVPLSFARQLWFLSQLE 4260
QY 1126 VGSWLWYLIPIYAVRMGRPVNVVDALRRALAAFORHETLRTTFEDQDQGVQVIVHEKLSSEM 1185
DB 4261 GPSATYNIPVALRDLGGPLDRDALTAAALHDVVYVARHEVLRVFTVADGEPWQQLDDPPQSV 4320
QY 1186 KVIDCGSDLDPEEVLNQEQTTPENLSSEAGMRATLRLGEDDHLTLTVMHHIISDGWSI 1245
DB 4321 PVVEVTFDRLP--FAVAAAGHRFDLGRELPRLAVLATGDDVHVVLVHVHIAAGWSM 4378
QY 1246 DVLRRDLNQLYSAALKSKDPLSALTPLPIQYSYFAKWKQDQFTEQ-----EKQLNRY 1298
DB 4379 RPLARDLAAVARI-DATAP--ALGALPVQYADYALWQDVLGSEHPDPSVISQVAYW 4435
QY 1299 KQKQKD--SSPAKIPDTFARFALLSGDAGCVHVTIDGELYQSLRAFPCNHNHTTSFVLLAA 1357
DB 4436 RRLAGVPEELDLDPVDRARPAEASHRGHTVEFAVPPAVHHQALAEARRNGVTVEMTVQTA 4495
QY 1358 FRAAHYRLTAVEDAVIGTPIANRNRPEDLIGCFVNTQCMRINIDHHDTEGTLINQVA 1417
DB 4496 LAVLLSKLGAGTDPIGVAVAGRTDPTLDNIGFFVNLVLTDLTGNTTITDLLHTRD 4555
QY 1418 TTTAAFENEDIPFERYVSALQPSGRDLSSTPLAQLIFAVHS--OKDLGRFKFGLESVPVP 1476
DB 4556 TTLHAFTHQDVPFEKLVEDLAP--TRSLARHPLFQVMMTLQASADEEPALAGLRVTDLP 4614
QY 1477 ---SKATYRDMERHFLQETDLSKGSVNFADLEPKMETVENVVRVFEILRNGLOSRTP 1533
DB 4615 AGETPAKVLDLTLHEVAGRDMHATLLGAADLEFEQETVRALAD---RLLTLEMAAAP 4671
QY 1534 VSILPLTDGIVTLEKLDVNL-----VKHVDYPR---ESSLADVQFQVSAVPSDSLAV 1582
DB 4672 DB-----RLDRIEVLSGERSRLLENWNTARPVVSSVPALFAEQVAAAPDAVAV 4722
QY 1583 VDSCLRTYTELDRQSDILAGWLRRRSMPAETLVAVFAPRSCETIVAFFGVILKANLAYLP 1642
DB 4723 VGEVSWTYRELDARSALRSVAAGVGESPVVVALERSPEVLSAFLAVAKAGGVFP 4782
QY 1643 LDVRSRQVQDILSG-----LSGPTIVL--IGHDTAPDPIEVTVNEFVRIDA 1689
DB 4783 VDLSPQARVDVAVVAGCARIADRPMSGLTVVSAGLGDSAV-----V 4827
QY 1690 LNDSDADGFEVIEHDSHKPSATSLAYLYTSGTGRPKGVMIHRVIRTVTSGCIPNYP 1749
DB 4828 SGDLTADRAVVL---PAGVPVGAAYRYMTYSGTGRPKGVVTHQNLVDLATDTCWGTP 4884
QY 1750 SETRMAMATIAFDGASYEIIYSALLFGRTLVVDYMTTLDARALKVFFREHVNAASHVT 1809
DB 4885 ---RVLPFHAPHAFDASSYEIWPVLLNGGTVV-VAPRPSIDATVLRDL-IGAHELTHVHT 4939
QY 1810 SSSQDVPVLRVPRRLSRTL--MFFFLVYVTDSTAPDALDAQLYQ-----GVQCXNGVGP 1861
DB 4940 AG-----LLRVLDPSCFAGLTVLTGGDAVSAEAVRRVKDANPGLRVRLYGP 4988
QY 1862 ENGVMS 1868
DB 4989 EVTLCA 4995

RESULT 6

US-09-815-242-5654
Sequence 5654, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5654
LENGTH: 1668
TYPE: PR1

ORGANISM: Staphylococcus aureus
US-09-815-242-5654

Query Match 8.5%; Score 1374; DB 10; Length 1668;
Best Local Similarity 27.3%; Pred. No. 7.9e-91;
Matches 459; Conservative 292; Mismatches 683; Indels 246; Gaps 62;
QY 461 SSMAEVNLMTEYDRAIESWNSQPLEVOD-----TLIHREMLKAVSHSPTKTAIOAW 512
DB 37 TTVDENLMTED---IQLYDDINSLPLDDAQVITLFEQO-----VEATFNHVAQ-F 88
QY 513 DGDW-TYSELNDVSRSLAVHIKSLGRLAQQAIIIP-----VYPEKSKWVIAASLAVLKSGN 566
DB 89 DGVFTYQTLNARANDLAHR-----LRNQGVPEPNDRAVIAEKSSTEMIAMIIGVLKAGG 143
QY 567 AFLTIDPNDPPARTAAVQTOTRATVALTSKLHRETVOKLVGRGVVVDDELLOSVSAS--D 624
DB 144 AVTIDPNTPSRQEQYILKDATPKVITYQALYENSKONINH---ID---LNKIANKNID 197
QY 625 DFSLSKSDLAIVFTSGTDPKGMIEHRAFSKALCFGASLGINSRDLALQFGTHA 684
DB 198 NLSECTLEDHAYVIYTSCTTGNPKGTLIPHRGIVRLVHQ-NHYVPLNEETILLSGTIA 256
QY 685 FGACILLEMTTLINGCGVCIPISSDDRMN--STPSFINRYNVMMATPSYMGTFSPEDVP 742
DB 257 FDAATFEIYGALLNGKLIIVAKKEOLLNPALAEQLINENDVNTMLTSSLFNQIASERIE 316
QY 743 GLATL-VLAVGOMSSVNAIAPKL-----OLLNGYGOSESSICFASNMSTE-PNN 792
DB 317 VLVPKYLIGGEV---LNAKWVDLLNQRKPKQIINGYGPTENFTTFTTYINPKVPNR 373
QY 793 --MGRV-GAHSWIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPPKSPFFTDIPS 849
DB 374 IPGKPIGLTHVIMQ---ERRCGVIGPGLCTSGFLAAGYLNQPELTADKFKID--- 427
QY 850 WYPANTFPDGAKYRTGDLARYASDGSIVCLGRIDPSQVKIRQORVELGAIEHLRQOMPD 909
DB 428 ---SNI---NQLMYRSGDIVRLPDGNDIDLYRKDKQVKIRGFRIELSEVE----- 472
QY 910 DLTIIVVEATKRSQSANSTSLIA-----FLIGSSYFGNRPDSAILHDHATKAINIKLE 962
DB 473 -----HALERIQGINKAVVIQNHDDQOYIV--AYY-----EAMHTLSHNKIKS---QLR 517
QY 963 QVLPKHSIPSPICMLPRTATGKIDRRRLRMCKDILDKOTGAIVQOAPAPIPVFPAD 1022
DB 518 MTLPEYMPVNFVHIEQIPITTINGKLDKALPIM--DYVD-----TDAYVAPSTD 565

QY 1023 TAAKLHSIWQSLGIDPATVNVGATFFELGGSNITAIKMN-MARSVGMDLKVSNIYQHP 1081
DB 566 TEHLQCQIFADILHVN--QVGIHNDFFELGGHSLKATLVNRIEATGKRLQIGOLLQKP 623
QY 1082 TLAGIS-AVVKGDPLSYTLIPKSTHEGVPQSQRLWFLDQDVGSLWLYLIPAVRMR 1140
DB 624 TVFELAQAIAKVOEQNEIPEIATVKKDDVYLSAOKRMVLLWKSNNHKDFVYVNPFLWRLS 683
QY 1141 GPVNVDAALRAALAELQRHETLRTTF-EDQDQGVGVQIVHEKLSSEMVKVIDLGCSDLDPE 1199
DB 684 SELNVAQLRQAVOHLIARHEILRTQYIVVDVQRQIVADVAD-----FE 729
QY 1200 VLNQEQT-----TPFNLSSEAGWRATLLRGEDEHILITVHHHIIISGWSIDVLR 1249
DB 730 EYNTHTDEQETMRQFVAPFNFLEKPSQIRVYIRSPHAY-LFIDTHHIIINDGMSNIQLM 788
QY 1250 RDLNOLYSAALKDSKPLSALTPLOYSDFAKWKQDOFIEQEKOLNWKWKOLKSSP-A 1308
DB 789 NDLNALLYQHKL-----LPLKLQYKDYSEWMSHRDWTKHO--YWLSQFKDEVPIL 837
QY 1309 KIPTDFARPALLSGDAGCVHVITIDGELYQSLRAFCEHNTTTFVVLAAAFRAAHYRLTAV 1368
DB 838 SLPTDYVRENIKTNGAMSFMTNQTRQLLQKYVEKHQITDFMFMSVVMTLLSRYARK 897
QY 1369 EDVICTPTANRNREPELIDICFVNTQCMRINIDHHDFTFGLINQVKATTTAAFNEDI 1428
DB 898 DDVVGVSVMSARMHKGAEQMLGMFANTLVYRGQSPDKMWTQFLQEVKEMSLAEYBHOY 957
QY 1429 PFERVVSALQPSRDLSSTPLAQLIFAVHSQK---DLGRFKFGLESVVPVSKAYTRD 1484
DB 958 PEECLVNDLQ-SHDASRNPLFDVMLVQNNETNIAHFGSKLTHIQ-----PKSVTAKED 1012
QY 1485 MEFLHFOEDSLKSGVSADELFKMETVENVYVVFELIRNGLOSSTRTVPSILPTDGTIV 1544
DB 1013 LSFIIEEDRDDYITINLEYNTDLYHSETVRHMGNCM-IMIDYLLKHQDLOICDIPNG-- 1069
QY 1545 TLEKLDVLN-----KHVDYPPRESSLADVFQTVQSAVPSLAVVDSRCRTYTELDRQSDI 1600
DB 1070 TEELLNVWTHVNDRLNMLNVPNGKSIISYFNEVYRQGNHVALVNDLTTIETLRYNYDA 1129
QY 1601 LAGWLRRRMPAETLVAFAFAPRSCETIVAFGLKANLAYLPDVRSPSARQDILSLGLS 1660
DB 1130 IAHILLNSGVNGQORVALFTERSFEMIAAMLATVKGASYIPIDIDFPKROGAILEDRAK 1189
QY 1661 GPTIILIG--HDTAPDPIEVNVE-FVRIR-----DALNDSNADGEFVIEHSDTKPSATS 1712
DB 1190 VTAVMSYGIETITLPIQLENAKGFVESKENEQYDDLHGDQLENTAMLDNE----- 1241
QY 1713 LAYVLTSGSTGRPKGVMIHVRVIRTVTSGCIPNYPSETRMA-----HMTATPADGAS 1766
DB 1242 -MYAIVTSGTTGMPKGVAIQRNLLMLVHA-----WSTELQGLDNEVFLQHANIVPDASV 1295
QY 1767 YEYSALLPGRTLVLCVDVYMTTLDAKALKDVEFREHVNAAASHVTSSSODVPLRVPRLSRT 1826
DB 1296 MEIYCCLLNGHILVIPDREERNVPAOLOQLINKHRVTAS-----IPLQMCV 1343
QY 1827 LMFFFL--VYDSTAPDALDAQLYQGVQC---YNGYGTENGVMSTIYPIDSTESFING 1881
DB 1344 MEDFYEKILITGATSTASEVK-YIEKHCGYFNAYGPSSEVITSYSHOCCGDLIPET 1401
QY 1882 VPIGRALNNGAYVVDPEQOOLVIGVMGELVTVGDGLARGY-----SDKALDENRFVH 1934
DB 1402 IPGKPLSNIQVYINS-EGILCGIGMPGELCIAGDSLTAIGYINRPELMADK-WQNNPF-- 1457
QY 1935 ITVNDQTVKAYTGRVRYRIGDGLIEFFGRMDTQFKIRGNRTESAEIEAALLRDSVDRD 1994
DB 1458 -----GKGLYHSGDGLARY-TSDGQLEFLGRIDQVKVNGYRIELDEIENAILAIGISD 1511
QY 1995 AAVVLQONEDQAEIIGLVFVVDHSDHSENDKGSANOVQWQGHFEGMYSIDIGEIDPSPI 2054
DB 1512 CVVTVSHFETH--DILNAYYVGEQVQEDLKQYLN-----DHLPKYMI-----PKTI 1556


```

; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
; FILE REFERENCE: 3019-PCT
; CURRENT APPLICATION NUMBER: US/09/976,059
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 1051
; TYPE: PRT
; ORGANISM: Actinoplanes sp.
; NAME/KEY: misc_feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: V represents a non-standard codon. It is expected that the biosyn
; OTHER INFORMATION: nthesized protein will have a formylmethionine residue at this po
; OTHER INFORMATION: sition
; US-09-976-059-13

Query Match          3.9%; Score 623; DB 9; Length 1051;
Best Local Similarity 21.4%; Pred. No. 2.6e-36;
Matches 328; Conservative 157; Mismatches 477; Indels 568; Gaps 42;

Qy 1114 SGRFLNFDLDVGLWLYLIPYAVRMKGPVNVDAALRRALAEQRHETLRTTFEDDQGVG 1173
Db 12 AQLAVHQAOLAPHSPIYQGEFEVDGECDDLLVAALQVMGEAESARLRFRVIDGTP 71
Qy 1174 VQIVHEKLSSEMVIDLCSGLDPFEV---LNQEQTTFPNLSSEAGWRATLRLGEDDH 1229
Db 72 WQYVAEDGDPIQVVDL-GAAADPRAAALGRMAADLRDGLRDPGLVEHHVLLGEGRV 130
Qy 1230 ILTVMHHSIDGWSIDVLRDLNQLYSALKSKDPLSALTPLPIQYSDFAKWKDQFI 1289
Db 131 IWYHRAHHIVCDGSLGIVASRVAGYSALAGDGVRRPGALPLPLSVLLSAADAYERSG-- 188
Qy 1290 EOEKQLNWKQKLDSSPAKIPDFAFPALLSG-----DAGCVHVTIDGELYQ 1337
Db 189 DRDRDRHWRSA-LA-GLPAELLAGACRPLPGPPVRRHEHDLAAAG----- 235
Qy 1338 SLRACNEHNTTSFVLLAAFAAHYRLTAVEDAVTGTPIANR-NRPELEDIIGCFVNTQ 1396
Db 236 RLKAGARRLTSVAQAGIAAALYQHRLTGARDVLVAVVAGRTTRPEF-DVPGMTSNV 294
Qy 1397 CMRINDHHDFTGLINQYKATTAFAENEDIFERVSVLQPGSDLSSTPLAQLIFAV 1456
Db 295 PVRLAYTPATTVGELLRDVARGVRDGLRHQRYPPNIVD-----DLGLADRAALRPVT 347
Qy 1457 HSQKDLGR-FKF-----QGLSEVPVPKAYTRFDMERHFLFOETDSLKGSVNFADLEFK 1508
Db 348 VNALALGRPLRFGSAGVRSGLSAGPVD-----DVTGLYKVSQ--GGMTIAELNP 398
Qy 1509 MET-----VENVVRVFEILRNGLQSSRT--PVSILPLTDGIVTLEKLDVLRKHVDYPRE 1562
Db 399 GRTDRPDAAEVSRWFTLLR-GLAESDAGDPVARIDIVDEPERRLLDENNATAA--PSS 455
Qy 1563 SSLADVFOQVSYAPDSLAVVSDSSRLTYTELDRQSDILAGWLRRRSMPAETLVAVFAPR 1622
Db 456 DTVLARFEEQAARTPEAPAVVCGDVTVTYAEAGANRLARVLARGAGPESVVALCLPR 515
Qy 1623 SCTIVAFAGVLKANLAYLPDVRKSPSARVODILSGLSGPTIVLIGHDT--APPDIEVTN 1680
Db 516 GPEVVTGILAAWAGAAAYLPDVTDELPAERVAYLL-GDSAAAVRLGTAETLAALPDGPAAD 574
Qy 1681 VEFVRIRDALNSDAGFEVIEHDSKPSATSLAYVLYTSGTGRPKGVMIHRVIIRT 1740
Db 575 VD-VHAPETIARESPP-----LRLEPLDQLAYIYTSSTGLSKGVGVSH----- 619
Qy 1741 TSCICINPSETMAHMAHTAFDGCAYEISALLFGRT-----LVCVDYMTLDA--- 1790
Db 620 --GGLANY-----VGMASLYGLSLAPLHSSLAFLDTVTSVFVPLVCGSVGVVSAAGG 671
Qy 1791 RALKDVFREHVNAAHSVTSQSDVP--LRVPRRLSRTLMEFFFLVVTDSSTAPDALDAQG- 1847
Db 672 RGLASLL-----AAGDGLSVKVPVGHLLR-----LLAELVPAGEMAAGVSLVAGGE 717

```

```

Qy 1848 -----LYQGVQCVNGYGPTEGVMSTIYPIDSTESFINGVPIGRALNNSGAYV 1895
Db 718 VLAGGDVRELSRVPGSVVNVGPTETVVGCSVFSVAAGDVVDVVPVGRVANTRLFV 777
Qy 1896 VDPEQQLVGIGVMGELVVTGDLARGYSKKA-IDENRFV--HITVNDQTVKATYRDRVR 1952
Db 778 LDEGLRPVPAGVAGELYVAGSQVARGVYGRSGLTASRFVACPFVGGE---RMYRTGDVVR 834
Qy 1953 YRIGDGLIEFFGRMDTQFKIRGNRIESAEATEALLRDSVSRDAAVVLQONEDQAPETLGF 2012
Db 835 LAGGD--LVFVGRVDEQVKIRYVE-----PDEVRLL 864
Qy 2013 VVADHDHSEKDKQGSANQVEGWDFHESGMYSDIGEIDPSTIGDFKGMTSMYDQSIDF 2072
Db 865 VVAGH----- 869
Qy 2073 DEMHEWLGETTRTLHDNRSLGNVLEIGTSGMILFNLDLSRLSESVVGLPSRRAAFVNSKA 2132
Db 870 ----- 869
Qy 2133 TESIPSLAGKAKVQVGTATDIGOVDDLHPDLVVLNSVIOYFSPSEYLAETADTLHLPNV 2192
Db 870 ---PRVAGAAYV-----ARPDV----- 884
Qy 2193 QRIFPGDVRSQATNEHFLLAARAIHTLGKNATKDDVRQKMAELEDMEBELLEVPAFTSLK 2252
Db 885 ----- 884
Qy 2253 DRFPGLVHEVILPKMNEAVNELSARYAAVHVHVRGSLGDELVLVPEKDDWIDFQANQLN 2312
Db 885 ----- 884
Qy 2313 QKSLGDLKSSDAAIMAVSKIPEITAFERQVVASLNSNIDENQLSIRSSAEGDSSLSV 2372
Db 885 -----GERQLVAYVVA----- 895
Qy 2373 PDIFRIAGE-AGFRVEYSSARQWSQNGALDAVFHCCSQGRTLVNFPTDHLRGSLLTN 2431
Db 896 -----AGEPAG----- 901
Qy 2432 RPLQRONRRIATEVRERLSLLPSYMIPSNIVVLDKMLNANGKVDKELSRRAKVVPK 2491
Db 902 -----LAESVRAHVAERLPEYMPVPAAVTLDIPTVNGKVDRAALPE-----PG 946
Qy 2492 QQTAAALPTFPISEVEVILCEATEVEFGM-KVDITDHFNFGLGHSLLATKLISRIDQRLK 2550
Db 947 PVATGNADREPTTERESLLCGAFADVLCIERVGVDDDFSLGGLHSLLATSLVSRVRLVIG 1006
Qy 2551 VRITVKDVPDHPVFADIASVIRQGLGQQP 2580
Db 1007 EELPIELFATPTPAELAAWLQNRADRPQ 1036

```

```

RESULT 9
US-10-166-087-44
; Sequence 44, Application US/10166087
; Publication No. US2003007767A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc.
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthramycin
; FILE REFERENCE: 3014-2US
; CURRENT APPLICATION NUMBER: US/10/166,087
; CURRENT FILING DATE: 2002-06-11
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 1446
; TYPE: PRT
; ORGANISM: Streptomyces refuineus subspecies thermotolerans

```



```

QY 1790 ARALKDVFFREHNAASHVTS-----SSQDVLPRVPRRLSTLMEFFLLVVD 1836
Db 236 PEALYQIEKEGVTVLSLTPSLMLMPARTANDKDL-----SLRTLFGGALSP 289
QY 1837 STAPALDAQGLYQGVOCYNGVPTENGWMSIYPIIDSTESFINGVPIGRALNNSGAYV 1896
Db 290 SLVDKWRERFGEKG-RLINTYGTPTTIVTTVNRITPEEIROKSVPIGRPLPNTQVYL 348
QY 1897 DPEQQLVGIGVAGELVVTC-DGLARGYSDK-ALDENRFVHITV-----NDQTVKAYRTG 1948
Db 349 DENGQLVPIGVAGELYIGWPGVARGYLNRPDLTAERFVNPFPQCEERGRNRRMYRTG 408
QY 1949 DRYVRIGDGLIEFFGRMDTPKIRGNRIESAIEAA--LLRDSVVRDAV 1997
Db 409 DLARW-LPDGTIEYLGRIDDQVKIRGYRIELGEIEAALLQHPGVKEAV 458

RESULT 14
US-09-994-595-8
; Sequence 8, Application US/09994595
; Publication No. US20030039981A1
; GENERAL INFORMATION:
; APPLICANT: Bhattacherjee, J.
; APPLICANT: Suvarna, Kalavati
; APPLICANT: Bhattacharjee, Vasker
; TITLE OF INVENTION: METHODS AND REAGENTS FOR DETECTING FUNGAL PATHOGENS IN
; FILE REFERENCE: 96,247-A
; CURRENT APPLICATION NUMBER: US/09/994,595
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 08/650,809
; PRIOR FILING DATE: 1997-05-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Microsoft Word 97
; SEQ ID NO 8
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-994-595-8

Query Match
Best Local Similarity 2.7%; Score 434.5; DB 9; Length 1391;
Matches 229; Conservative 206; Mismatches 402; Indels 321; Gaps 49;

QY 233 THFWOTHLNDLASVFPPLSHDLMPNPTTAEHRITFPLSOKALSNSAICRTALSI--- 289
Db 2 TDFWNLNDLPNTLSVLPH---DFLKPANNKSVEGTYFNID---NGSTDFFGLAVFAA 54
QY 290 LLSRTHSDEALFAGVTEQSLPDKHYLDGTYQTVAPLRVHCQSNLRASDVMDALS-SY 348
Db 55 LVYRLT-GDEDIVATDESA-----NTPEFI--VRLNLTPELTPELVSKITKEY 101
QY 349 DORLGLHAPFLGRDIRNTGDNGSAACDFQVLLVLTGDSHVNNGINGFLOQITESSHFPC 408
Db 102 ENNISQI-----NYKALSEVSHRKEAKGLDENPCLFRLSYQHAHNSOOLNT 148
QY 409 NNRALLHCOMESSGALLVAYYDHNVIDSLQTLRLLQFQHLKICLQSPDLSS----- 462
Db 149 TVEGSIRDLAIVTDGTFKFTIYN-----ALLYSHERVVIGCEQFAQLTTVSGDDT 198
QY 463 --MAEVLNMTEDRAIE-----SWNSQPLEVQDQTLIH-----MLKAVS---HSP 504
Db 199 TVIAEVFLTDFPHKMLNPDTDLQMSGYRGAIQEIFMDNANKHPDQTCVETVPSLESN 258
QY 505 TKTAIQAWGDWTSYSELQVSSRLAVHKSGLRAQQAIIPIVFEKSKVVIASMLAVLKS 564
Db 259 SKTR-----NFSYHKLKLLIVGNLYKETGK-KGDIVMAYRGVDMTAVMGVLKA 311
QY 565 GNAFTLIDNPDPAR-----TAQVVTQTRATVALTSKLHR 599
Db 312 GATFSVIDPAYPARQNIYLVAKPKGLIGLEKAGTLDQLVDVYISNELDVVSTIPQLKV 371
QY 600 ETQVKLV-GRCVVVDDELQSVASDDFSLSLTKSQDLAYVI-----FTSGSTGDPK 649

```

```

Db 372 QDDGTLVGKLEGANDCL-----NDYQKF-KDQPAVIVGPDSTRPTLSFTSGSGEIPK 424
QY 650 GIMIEHRA-----FSSCALKFGASLGINSOTRALQFQTHAFGACLLIEMITTLINGCVCIP 705
Db 425 GVLGRHYSLAYFFPMMAKRFLS---EKDKFTLSGI-AHDPIQORDMFTPLFLUGAQLLVP 480
QY 706 SDD--RMSIPSFINRYN---VNMMAATPSYMGTFSPEDVPGLATLVLVGEOMSSS--- 757
Db 481 TADDICTPGKLDWMAKYGATVTHLTLMGQLLSAQAATAIPSLHAF-FVGDILTTRDCL 539
QY 758 -----VNAIWA-PKLQLLNGYQ--SESSSICFASNSTPNMNGRAVGAHS--- 801
Db 540 RLQSLAENFIVNMLWSLSQTSQSVSYFEIKSRKADPTYLKNLKAV---MPAGTGMHVQ 596
QY 802 -WVIDNDINRLVPIGAVGELVIESPGIARDYIVPPPEKSPFTDIPSWY----- 851
Db 597 LLVNRNDRSQTGCGVEGEIYVRAAGLAEGYRGLDPLNAKEFIT---NMVNPDKWIEQ 653
QY 852 -PAN-----TFPDGAKLYRTGDLARYASDGSIVCLGRIDRISQVIRGQVRELGA 899
Db 654 DEANKKSETSETRTWSVKPRDRMYRSGDLGRYFSDGNECCGRADDQVKIRGFRIELGEI 713
QY 900 ETHLROQ--MPDOLTIIVTEATKRSQSANSTSLIAFLIG-----SSYFGN-----RPSDA 946
Db 714 DTHLSQHPLVRENVTLV-----RRDKNEEPTLISYIVPKDSPELKFTFFADVDPPLKSN 768
QY 947 HILD-----HDATKAINIKLEQVLPRHSIPSPFICMLELPRTATGKIDRRRLRIMGK 1001
Db 769 PIVKGLVAYRELKIDKYLKKLASAIPITIIIVPLVKLPNPNKGVDPKPL----- 820
QY 1002 DKQTQGAIVQQAIPAPIPVFAADTA-----PPFDTAQLAAVAKLSVSSHDAQAABEENLT 1037
Db 821 -----PPFDTAQLAAVAKLSVSSHDAQAABEENLTKEEQIRDLWLDVLPN 866
QY 1038 DPATVNVGATFFELGNSITAIKMVMNARSVMGLKVSNIYQIHTLAGISAVVKGD----- 1093
Db 867 RPATISKDDSFEDLGSLSILGTRIFTYEOKLNVIEPLVSP-----KGDQRRP 913
QY 1094 --PL-----SYTL-----IPKSTHEGPVEQSYSGRLWELDQDLD 1125
Db 914 RPIGLSRVNYSRREORRRFLKAKTYTMRRSKELSKLSKALSSEYSS-----LKQLP 968
QY 1126 VGSWLYLIPYAVMRGPNVNDALRALAALAEQRHETL-----RTTFED-----QD 1170
Db 969 SGSVNVFVGTATGFLGSFIVRDL-----LTARNKNLDIKVYAHVRASSKEAGLQRLQT 1022
QY 1171 GVGQVIVHEKLSSEMKVI 1188
Db 1023 GITYGIWDENWAEKIEIV 1040

RESULT 15
US-09-738-626-6480
; Sequence 6480, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16

```

```

: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 6480
: LENGTH: 1295
: TYPE: PRT
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-6480

```

| Query Match | 2.6% | Score 424 | DB 9 | Length 1295 |
|-----------------------|------------------|---|------------|-------------|
| Best Local Similarity | 25.8% | Pred. No. 1.3e+21 | | |
| Matches 159 | Conservative 112 | Mismatches 243 | Indels 102 | Gaps 25 |
| QY | 495 | EMLKAVSHS-PTKTAIQMGD-WTYSSELDVSSRLAVHKISGLRAQQAIIIPVYEKSK | 552 | |
| Db | 24 | DVLESVASTYPPAAALD--DGQVLTVAELMEEVTALADSIHAGIIRGRDGRIGTMRPSGTR | 81 | |
| QY | 553 | WVIASMLAVLKSGNAFTLIDPNPPARTAQVVTQTRATVALTSKLHRETQVKLVGRGVV | 612 | |
| Db | 82 | DLYTIALLATLAAGAAVVPVDDADPEERAEMVFEANINALFDA-----T | 125 | |
| QY | 613 | DELLQSVSASDDFSLTSKQSLAYVIFTSGSTGDPKGMIEHR-----AFSSCALKF | 665 | |
| Db | 126 | GPHMLRPTAGGB--TRRRLDDTATFTFSGSTGKPGKGVASHRSAAAFVDAQAQPLVD | 183 | |
| QY | 666 | GASLGINSRALQFQTHAFGACLEIMTTLINGGCVCIPISDDDRMNS---IPSFINRYN | 722 | |
| Db | 184 | HPSGPLGPDRLVAGLSVAFDASCEEMLAWGHGACL-VPAPRSLVRSGLMDLGPWLIRRD | 242 | |
| QY | 723 | VNMNMATPSYMGTFSPEDYVGLATLVLGEQMSSS-VNAIAPKLOLLNQVSGOESSI- | 780 | |
| Db | 243 | ISVSTVPPLAGLWPAEALSRVLLVGVGACSQELVERLSTPDRVWNTYGTGTEATVA | 302 | |
| QY | 781 | CFASNKSTENPNKRAVGAHSHWIDPNIDNRLVPIGAVGELVIESPGIARDYIVPPPEPK | 840 | |
| Db | 303 | CGTQLYAGQPVGIGLPLAGMDLVV-VDDAGEPVGIGEVGELVIGGVGLAR-YL--DPEK | 357 | |
| QY | 841 | SPFFTFDIPSWYFANTFPDGAKILYRGGDLARYASDGSIVCLGRIDSQVKIRQVRELGAIE | 900 | |
| Db | 358 | -----DREKAPLKSV-GWTRAYRSGDHLVLEEDG-LYFVGRVDDQGVKGGRIELGEVD | 410 | |
| QY | 901 | -----THLRQOMPDLTTVWEATRKSQSANSTSLIAFLIGTSSYFGFNRPSSDAHILDHAT | 954 | |
| Db | 411 | ANVAALSNVRS-----AVVQTT-----GADQKVLVAVV-----SLEDAAGFDHNV | 454 | |
| QY | 955 | KAINTKLEQVLRPHSIPSYICMLELPRATGCKIDRRRLRMINGKOLDKQTQCAIVQOAP | 1014 | |
| Db | 455 | TA-----RLTETMPAALVPRITHV-MDDLVPVTTSGKVDRKSL-----P | 490 | |
| QY | 1015 | APIP-----VFADTAALKHSITWQSLGIDPATVNVGATFFELGNSNTAIKWNMARS | 1067 | |
| Db | 491 | WPLPGTVVEANDLSATEAWIAQEWVDILGTSVSSKD--ADFFSLGCTSAAATLVGRVRA | 548 | |
| QY | 1068 | VGMDLKVSNIXQHPTL 1083 | | |
| Db | 549 | KVPTAAVRDLYDHPRL 564 | | |

Search completed: May 30, 2003, 13:01:36
Job time : 105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:33:51 ; Search time 39 Seconds
(without alignments)
3327.678 Million cell updates/sec

Title: US-09-482-788-2

Perfect score: 16128

Sequence: 1 MEXLTAVDGRDLPPTASF.....RVEHLEEVSKTFEGLNSSL 3129

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|--------|-------|--------|--------------|---------------------|
| 1 | 2084 | 12.9 | 6486 | 1 TYCC_BACBR | O30409 b tyrocidin |
| 2 | 1975 | 12.2 | 6359 | 1 BACC_BACLI | O68008 b bacitraci |
| 3 | 1882.5 | 11.7 | 5255 | 1 BACA_BACLI | O68006 b bacitraci |
| 4 | 1816 | 11.3 | 3587 | 1 TYCB_BACBR | O30408 b tyrocidin |
| 5 | 1807.5 | 11.2 | 2555 | 1 PPS3_BACSU | P39847 bacillus su |
| 6 | 1793.5 | 11.1 | 3587 | 1 SRF2_BACSU | Q04747 bacillus su |
| 7 | 1769.5 | 11.0 | 2560 | 1 PPS2_BACSU | P39846 bacillus su |
| 8 | 1758 | 10.9 | 4451 | 1 GRSB_BACBR | P14688 b gramicidi |
| 9 | 1710 | 10.6 | 2607 | 1 BACB_BACLI | O68007 bacillus li |
| 10 | 1625.5 | 10.1 | 5217 | 1 HTSL_COCCA | Q01886 cocciliobolu |
| 11 | 1613 | 10.0 | 3587 | 1 SRF1_BACSU | P27206 bacillus su |
| 12 | 1604 | 9.9 | 2561 | 1 PPS1_BACSU | P39845 bacillus su |
| 13 | 1515 | 9.4 | 3712 | 1 ACVS_CEPAC | P25464 cephalospor |
| 14 | 1484.5 | 9.2 | 3649 | 1 ACVS_NOCLA | P27743 nocardia la |
| 15 | 1422.5 | 8.8 | 3770 | 1 ACVS_EMENI | P27742 emericella |
| 16 | 1347.5 | 8.4 | 3791 | 1 ACVT_PENCH | P26046 penicillium |
| 17 | 1346.5 | 8.3 | 3746 | 1 ACVS_PENCH | P19787 penicillium |
| 18 | 958 | 5.9 | 3947 | 1 STD2_USTMA | O43103 ustilago ma |
| 19 | 909.5 | 5.6 | 1278 | 1 DBHF_BACSU | P45745 bacillus su |
| 20 | 852 | 5.3 | 4447 | 1 PKSJ_BACSU | P40803 bacillus su |
| 21 | 756 | 4.7 | 1274 | 1 SRF3_BACSU | Q08787 bacillus su |
| 22 | 742 | 4.6 | 1098 | 1 GRSB_BACBR | P14687 bacillus br |
| 23 | 740 | 4.6 | 1088 | 1 TYCA_BACBR | P09095 bacillus br |
| 24 | 654.5 | 4.1 | 2035 | 1 HMP2_YEREN | P48633 yersinia br |
| 25 | 568 | 3.5 | 1293 | 1 ENTF_ECO57 | O8xbv9 escherichia |
| 26 | 559 | 3.5 | 1293 | 1 ENTF_ECOLI | P11454 escherichia |
| 27 | 514 | 3.2 | 1419 | 1 LYS2_SCHFO | P40976 schizosacch |
| 28 | 494.5 | 3.1 | 1392 | 1 LYS2_YEAST | P07702 saccharomyc |
| 29 | 469 | 2.9 | 805 | 1 ACVS_STRLC | Q01757 streptomyce |
| 30 | 451 | 2.8 | 1048 | 1 ANGR_VIBAN | P19828 vibrio angu |
| 31 | 434.5 | 2.7 | 1391 | 1 LYS2_CANAL | Q12572 candida alb |
| 32 | 408.5 | 2.5 | 503 | 1 DUTA_BACSU | P39581 bacillus su |
| 33 | 383 | 2.4 | 506 | 1 DUTA_LACCA | P35854 lactobacilli |

RESULT 1

| ID | TYCC_BACBR | STANDARD; | PRT; | 6486 AA. |
|----|--|-----------|------|----------|
| AC | O30409; | | | |
| DT | 15-JUL-1999 (Rel. 38, Created) | | | |
| DT | 15-JUL-1999 (Rel. 38, Last sequence update) | | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | | |
| DE | Tyrocidine synthetase III [Includes: ATP-dependent asparagine | | | |
| DE | adenylase (AsnA) (Asparagine activase); ATP-dependent glutamine | | | |
| DE | adenylase (GlnA) (Glutamine activase); ATP-dependent tyrosine | | | |
| DE | adenylase (TyrA) (Tyrosine activase); ATP-dependent valine adenylase | | | |
| DE | (ValA) (Valine activase); ATP-dependent ornithine adenylase (OrnA) | | | |
| DE | (Ornithine activase); ATP-dependent leucine adenylase (LeuA) (Leucine | | | |
| DE | activase)]. | | | |
| GN | TYCC. | | | |
| OS | Bacillus brevis. | | | |
| OC | Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus. | | | |
| OX | NCBI_TaxID=1393; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=ATCC 8185; | | | |
| RX | MEDLINE=98012987; PubMed=9352938; | | | |
| RA | Mootz H.D., Marahel M.A.; | | | |
| RT | "The tyrocidine biosynthesis operon of Bacillus brevis; Complete | | | |
| RT | nucleotide sequence and biochemical characterization of functional | | | |
| RT | internal adenylation domains."; | | | |
| RL | J. Bacteriol. 179:6843-6850(1997). | | | |
| CC | -!- FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN, | | | |
| CC | GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE | | | |
| CC | PEPTIDE PRODUCT. | | | |
| CC | -!- COFACTOR: CONTAINS 6 COVALENTLY BOUND PHOSPHOPANTHETINES (BY | | | |
| CC | SIMILARITY). | | | |
| CC | -!- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis. | | | |
| CC | -!- SUBUNIT: LARGE MULTIMERIC COMPLEX OF TYCA, TYCB AND TYCC. | | | |
| CC | -!- DOMAIN: CONSISTS OF SIX MODULES, AND HARBORS A PUTATIVE | | | |
| CC | THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE | | | |
| CC | INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE | | | |
| CC | FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE- | | | |
| CC | ADENYLATION, THIOLEATION, CONDENSATION (NOT FOR THE INITIATION | | | |
| CC | MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION | | | |
| CC | (OPTIONAL). | | | |
| CC | -!- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC | | | |
| CC | DECAPEPTIDES; TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL- | | | |
| CC | ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR | | | |
| CC | RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE | | | |
| CC | CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM. | | | |
| CC | -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME | | | |
| CC | FAMILY. | | | |
| CC | -!- SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |

or send an email to license@isb-sib.ch.

```

CC -----
DR EMBL; AF040835; AAC45930.1; -
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatin.
DR InterPro; IPR003880; Phantne.attach.
DR InterPro; IPR000379; Ser_estrs.site.
DR InterPro; IPR001031; Thioesterase.
DR pfam; PF00501; AMP-binding; 6.
DR pfam; PF00550; pp-binding; 6.
DR pfam; PF00668; Condensation; 6.
DR pfam; PF00975; Thioesterase; 1.
DR PRINIS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
DR PROSITE; PS00455; AMP_BINDING; 6.
DR PROSITE; PS00075; ACN_DOMAIN; 6.
KW Ligase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.
FT REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).
FT REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).
FT REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).
FT REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).
FT REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).
FT REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).
FT DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
FT DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.
FT DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.
FT DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.
FT DOMAIN 5124 5191 ACYL CARRIER (ACP) 5.
FT DOMAIN 6167 6234 ACYL CARRIER (ACP) 6.
FT BINDING 1000 1000 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2037 2037 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3075 3075 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4110 4110 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 5154 5154 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 6197 6197 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT SEQUENCE 6486 AA; 724011 MW; 4934900AF07DF786 CRC64;
Query Match 12.9%; Score 2084; DB 1; Length 6486;
Best Local Similarity 23.0%; Pred. No. 9.8e-112;
Matches 864; Conservative 595; Mismatches 1372; Indels 928; Gaps 136;
QY 46 IEAKPCTPQDMDICNALDKQSAIGHAYD-----VPTDIDISRFALANK 92
DB 7 IAKIYPLTLQ-----EGMLFHAVTDTGSSAYCLQMSATIEGDFHLPFEKSLN 55
QY 93 EIVNQTALRAFAFTSDGKTSQVILKDSFVFWMCWSSSSSDPEVVVDEAAAAAGPRC 152
DB 56 KLVENYELRTAFVYQNMQRPRQVVEKRV-TVPCENTAHLP--AEQDAYIQAYTKQH 112
QY 153 NRVLLD-----MGT--KKQQLVWTFSHALVDVTFQOVRVSRVFAAYKHEKDTHRPE 203
DB 113 HAPDLTKDNLKMAAIFQTAENKYRLVWAFHHIIVDGTGLVLLHKLTYVA-----ALRKG 168
QY 204 TPSSDATDTSQSVSWSCEDNAVSAFTHFWQTHLNDL-----ASVFPH 250
DB 169 EPIPREATPYSEYIKWLDKQNKDEALA---YQNYLAGYDHOAAPPKKKLGTEASRYEH 225
QY 251 LSDHLVMPNPTTAEHRITFPLSOKALSNAICRTALSTILLSRYTHSDRALFGAVTEQSL 310
DB 226 VEAMFTIAPEKTOQLQIA---NONOATSSVFOALWGLASTYKKNADDDVVFSGVSGRP 282
QY 311 P-FDKHYLAGDTQYTVAPLRVHCQSNLRASDMDAI-----SSYDRLGLHLPGLRD 362
DB 283 PQIOGIESWGLFINTIPTRVQTNKQOTSELLQTVQKALASATVD-----FAP--LYE 335
QY 363 IRNTGNGSAACDFQVL-----LVTGDSHVNNNGINGFLQOITESSHFH-----406
DB 336 IQST-----TVLKQELIDHLVTFENYPDHS-----MKHLEESLGFQFTVESGDEQ 380
QY 407 -----PCNNRALLLHCQMESSGALLVAYYDHNVIDSLQTTLLQOPGHILKCL 454

```

```

DB 381 TSYDLNVVVALAPSNE-----LYVKLSYNAAV-----YESSFVNRIE-----GHURTVI 424
QY 455 QSPDLDS--SMAEVNLMTEYDRAE-IESWNSQPLEV--QDTLIHHEMLKAVSHSPTKTAIQ 510
DB 425 DQVIGNPHVILHEIGIITEEEKQOLLVAYNDTAAEYPRDKTIFELTAEQASRTPAKAAVV 484
QY 511 AMDGWTYSELDQNVSSRLAVHIKSLGLRAQQAIIPIVYFEKSKWVIASMLAVLKGNAFTL 570
DB 485 CGEDTLTYQELMERSAQLANALREKGI-ASGSIVSIAEHSLELIIVAIMAVLRSGAAYLP 543
QY 571 IDPNPPARTAQVVVTRATVALTSKLHRETVOKLVGRVGVVDDDELQSVASDDSSSLT 630
DB 544 IDPEYQDRIQYLDDSQTTLLTQS-HLPNTRFAGSVLYLDDRSLYE--GGSTSFAPES 601
QY 631 KSQDLAYVFTSGTGDPRGIMIEHRAFSKALKFGASLGINDTRALQFGTH---AFGA 687
DB 602 KPDDLAYMIVTSGTGNPKGAMITHQGLVN-YIWANKVYVOGE--AVDFPLYSISFDL 658
QY 688 CLLEIMTTLINGCCVCIPISDDDRMNSIPSFIRYNNVNMWMTSPYMGTFSPEDVPLATL 747
DB 659 TVTSIFTPLLSGNTIHVYRGADKVQVILDIKDKNGVGIITLPTHLKLI--EHIDGKASS 716
QY 748 V---LVG-----BOMSSSVNAIWAPKLQLLNGYGQSESSICFA---SNNSTEPNNNGRA 796
DB 717 IRRFIVGGENLPTKLAKQIYDFEGENVQIFNEVGPTEYVVGCMYIYLDPTQTQESVPIG 776
QY 797 VGAHS---WVIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPEKSPFTDIPSWYPA 853
DB 777 VPADNVQVLLDAS--MQPVPVGLGEMYTAGDVAKGVFNRPELTKEKFI-----826
QY 854 NTFPDCAKLYRTGDLARYASDGSIVCLGRIDSQVIRGQVORVELGAIETHLRQ--OMPD DLT 912
DB 827 NPERPGTKMYRTGDLAKWLPDGNMEYAGRMQYQVIRGHRIEGELIETRLTQHEAVKEAV 886
QY 913 IVVEATKRSQSANSTSLIAFLIGSSYFGNRPNSDAHILDDHDKA--INIKEOVLPKHSIP 971
DB 887 VIVEKDESGQNV---LYAYLVS-----ERELTVAERREFLGRITLPSYMIIP 928
QY 972 SFYICMLELPRATGKIDRRRLRMKGIDLDKOTGAIVQOAPAPIPVFAADTAALKHSIW 1031
DB 929 SFPIRLAEIPLTANGKVERKKL-----PKPAGAVVTGTAYAAPO-NEIEAKLAEIW 978
QY 1032 VQSLGIDPATVNVGATFFELGNGSITAIKMW-NKARSVGNLDKLVSNYIHOPTLAGISAVV 1090
DB 979 QQVVLGI--SOVGTHDDFDLGGHSLKAMTVVFOVSKALEVELPVKALFEHPTVAELARFL 1036
QY 1091 -KGDPLSYTLI-PKSTHE-GPVEQSYSGRLWFLDQLDVGSLWYLIPIYAVMRGPNVDA 1147
DB 1037 SRSEKTEYTAIOPVAQEEYPV--SSAQKEMYILQOFEKGNGISYNTSAILLESGLDYAR 1094
QY 1148 LRRALALFORHETLRTTFEDQDQGVQIVHEKLSBEMKVIDLCGSDLDPEFVNLQEQTT 1207
DB 1095 FASAVQQLAERHEALRTSFHRIDGEPVQKVHEEVEVPLFMLE---APEQAEKIMREFVR 1151
QY 1208 PNLNLSAEGWRATLLRLGEDDHLITIVMHHIISDGSIDVLRRLDQLNQLYSAALKDSKDL 1267
DB 1152 PFDLGVAPLMRTGLLKLGRHLFLDMDHHIISDGVSSQILLREFAELOYGA-----1203
QY 1268 SALTPPIQVSDPFAKWQKQDFIEQ--EKOLNYYKKQLKDSPP-AKIPDFAFPALLSGDA 1324
DB 1204 -DLQPLISLQYKQFAANQNELFQTEAYKKQEQHMLNTFADEIPLNLNPTDYPFVSQVSFAG 1262
QY 1325 GCVVHTIDGELYQSLRAFCNEHNNTTSFVLLAAAFRAAHYRLTAVEDAVIGTPIANRRPE 1384
DB 1263 DLVLFAGKELLERLQOVASETGTTLYMILLAAYNVLLSKYTQGTQEDIIIVGTVPAGRSHAD 1322
QY 1385 LEDIICCFVNTQCMRINIDHHDFTGLINQVKATTTAAEFENEDIPERVVSALQPGSRDL 1444
DB 1323 VENIMGIFVNTALRNOPASSKTFQAFLOEVKONALAAVDHODYPPEELVEKLAII-ORDI 1381
QY 1445 SSTPLAQLIAFVHVSQDKLGRFKFGLESVPVPSKAY-----TRFQMEFHLFQETDLSKG 1498
DB 1382 SRNPLDFTLSLENANQ-----QSLAIAELTASPELVFNKISKFDLALNASESPADIQF 1435

```


496 QBARLLREKGI--GRSIAAIIADRSEFEMIIGIILKAGGAYLPIDPETPKHRIAPMLSD 554
 587 TRATVALTSKLHRETQVLGRVQVVDDELQSVASDDF-----SSLTKSODLAVYFT 641
 555 TKAGVLLAQGAADID-----CEA--DIIHLDKGADGFSKRLSSVNDSDTAYIYT 607
 642 SGSTGDPKGIEMTEHRAFSKALKFSGAGINSIDTRALFGTHAFAGACLEIMTTLINGCC 701
 608 SGSTGMPKGVVTPHYS--AARVVKNTNYIDITEDDAILQLSNYSFSGSVDFIDFALLNGAS 666
 702 VCIPSDDDRMNS--IPSEFINRYNVNMWATPSYMCCTFSPEDVPGLATL---VVLVGEQMS- 755
 667 LVLIKEVTLNTHAEVETKREQSVSMFTTALFNTLADININGCLAKURKIFLGGERASI 726
 756 ----SSVNAIWAIPKQLLNGYQSES---SSTCFASNMSTENNMGRAGV---AHSWVID 805
 727 PHVRKVLNVRGDK--LIHVGYTESTVYATVYFINEIDDEAETI--PIGSPPLANTSVL 782
 806 PNDINRLVPIGAVGELVETSPGIARDYI-----VPPPEKSPFTTDPSPWPANT 855
 783 MDEAGKLLPIGVPGELCIAGDLSKGYLNREBELTAKEFIPHP----- 824
 856 PDGAKLYRTGLARYASDVCLGRIDSQVKIRGORVELGAIETHLRQMPDDL--TI 913
 825 FTIPGERLKTGLAKWLDGNIEFIRIDHQVIRKGFRIELGEIESRL--EMHEDINETI 882
 914 VVEATKRSOSANSTSLIAFLIGSSYFGNRPDSAHILHDHATKAINIKLEQVLPRIHSPF 973
 883 V---TVREDEESRPYICAVIT-----ANRE-----ISDELKGF---LGEKLPYEMIPAY 926
 974 YICMLELPTATGKIDRRRLRMGKDILDKQTOGAIVQ--QAPAPIVFADTAAKLHSIW 1032
 927 FVKMDKLLPTKNGKVDKALP-----EPDRTAGAEYEAEP-----NETEKLAAIWR 975
 1033 QSLIGDIPATVNVGATFFELGGNSITAIRM--VNARSVGMDLKAVSNLYOHPITLAGISAVVK 1091
 976 DILKVEKSGIN--DIEFFEMGHSKAAAMAARIKELKAEIPLGQIFKPTTKGLEYIR 1033
 1092 GDPLS-YTLPIKSTHEGPVEQYSQGRWFLDOLDVGLWYLIPIYAVRMGRPNVDALRR 1150
 1034 STKDSYSSIQVEEKEYRLSSAOKRLYLIDQIEGSLSYNIPFTMKVKGRDIRRFEN 1093
 1151 ALAALEORHETLRTTFEDODGVGVQIVHEKLEEMKVIDLCSDDLDPFVLNQEQTTTPN 1210
 1094 ALKTIQRIHEALRTFLMADGEPVQIEKEVDFSKCKSI--QSLSIQIEIIRKQ-FVRPFD 1150
 1211 LSSEAGWATLLRLGEDDHLITVMHHIISDCWSIDVLRDLNOLYSAALKDKDPLSAL 1270
 1151 LKAPLFRTEVVKVDDDEHHIILFDMHHIISDASMGVLTKEICDLYG-----KEL 1201
 1271 TPLPIQYSDFAKQKD--QFIEQKQNLNWKQKLDSSPA-KIPTDFARPALLSGDAGCV 1327
 1202 KPLSLQYKDYSEWQDFYQKDEMKQKEYWLNIKGEIPLVNMPTDYPROMHSEBGDRI 1261
 1328 HVTIDGELYQSIRACNEHNTTSFVVLAAFAAHRYLTAVEDAVIGTPIANRNPELED 1387
 1262 GRAIDGELTKLKRIAKDNGATMYMLLAAYTVLLRTYSQOEVDVITGTIOGRKHHELKH 1321
 1388 IIGCFVNTQCMRINDHDDFTGLINOVKATTAAAFENEDIPFERVVSALQPCGRDLSST 1447
 1322 VLGMEVNTLAMRNHPKDKTFAYEIQDVKETALKAYENQDYQDFDLVEQLDL-ERDMSRN 1380
 1448 PLAQLIFAVHSOKDLGRKFQGLSEVPVPSKAY-TRFDMEFHFLQFTSDLSKGSVNPADL 1506
 1381 PLFDTMFVLQNLKADA-BIEGLTEPPFESDIHISKFDLTLSAIEKDSKIEFDLEYCTKL 1439
 1507 FMNETENVVRVFFELRNLGLOSRTPVSIPLDTGIVTL---EKLDVL---NVKHVDYVP 1560
 1440 FKRETVERMAAHFVRVLED--ISKRTDKRL---DQIEAMSEDEKNTLLYRFENDTKTDAP 1493
 1561 RESSLADVPQTOVSAYPDSLAVVSDSCRLTYTTELDRQSDILAGLRRRSMAPETLVAFA 1620
 1494 TDKTIQCLFAERAETSPDKTAVVFEDQTLTYRQLHRSNQLARFUREKGVQPDATVGINV 1553

1621 PRSCETIVAFFGVGLKANLAYLPDVRSPSARVQDILSGLSGPTIVLIGHDTAPPDIEVTN 1680
 1554 DRSPEMIGLLGILKAGGAYLPDPAYPEDRIKYILG--DSQTKFLLSEAL-----IKK 1606
 1681 VEFVRIRDALNDSNADGFEVIEHDSHTKPSATS-----LAVLYTSGTGPKGPMIHRVI 1736
 1607 RAFIREKADINDIIDK-QIAAQDAQAQLEFVPSRSGDLAYIIVTSGTGPKGVLIPQKGL 1665
 1737 IRTVTSGC-IPNVPSETRMAHMATIAFDGASVEIYSALLFGRTLVCDVYMTTLLDARALD 1795
 1666 SNLVSAYVKMLHNTGSRVIOFASUSFDASAEIFFALAGSALVLCROEEMPGQPLRS 1725
 1796 VFEFHVNAASHVTSSQDVPLRPRRLSRTLMFFVFLVYTDSTAPDALDAQGLVQGVQY 1855
 1726 -FLQY--NITHAT-----LPTVLDVNLSEGLNKLKIVSAGSACSEELAKRWSGNRLF 1777
 1856 -NGYPTENGVMST--IYPIDSTESFINGVP-IGRALNNGSAYVVDPEQLVGLVIGVMGEL 1911
 1778 INAYGPTETVTCATAGIYE-----GSRPHIGSPIANTNVYLDONKQKPTGVVYVGL 1830
 1912 VVTGDLARGYSDK-ALDENRFV-HITVNDQTVKAYRTGDRVRYRGDGLIEFFGRMDTQ 1969
 1831 CVGGSRLARGYLNRPPELTAEKFIHPFASGE--RLYRTGDLARW-LPDGHLEPLGRIDHQ 1887
 1970 FKIRGNRIESAEIEAALLRDSVYDAAVVLOQNEQAPFELGVFVADHDHSENDKQGSAN 2029
 1888 VKIRGYRIELGEIENOLLKLDKIDEAAVIAARKDDHSDVLCAYIVSKED----- 1936
 2030 QVEGWQDHFESGMYSDIGEIDPSTIGSDFKGWTSWYDQSGDIDEMHWEHGLGTRTLHDN 2089
 1937 -----WTS-----TEISEW----- 1945
 2090 RSLGNVLEITGCTGSMILNLSRLESYVGLFSPSAAAFVFNKATESIPSLAGKAKVQVGT 2149
 1946 -----LEKELPHYM-----IPAYFVR--LDKLP----- 1966
 2150 ATDIGOVDL-----HPDLVVLNSVIOYFSPSEYLAETADTLIHLPNVOR--IPFGDVRVSOA 2204
 1967 LTSNDKVDKRALPAPORHVATGAVYEAPRNDTEAKLVD-----IWRDVLGAGDI---G 2016
 2205 TNEHFLAA-----RAHTLGKNATKDDVRQKMA--ELEDMEBELLVPEAPFTSLKDR 2254
 2017 ISHFFAAGDSIKALQIVSRJSLGKLEMKDOLFANPRIKDLAKYVKKO-----SORKNA 2072
 2255 FPLGVHEVTELP-----KNMEAVNELSAYRYAAVHVVRGSLGDELVLPL-----VEKDD 2302
 2073 NTIVTGHAEALTPIQKWYFANNKE--ELDHFNQSFVLFRKGGFDESCVKKAFNKIMEQHD 2129
 2303 WI-----DP-QANQLNQKSLGDLKSSDAAIMAVSKIPPEITAFERQV--VASLNS 2350
 2130 ALRMIYEEKGGDFIQNRSFRDELFDL---DV-----YDVRGLDQAEKVVELAT 2176
 2351 NIDEWOLSTIRS-----SAGDSSLSVPDIFRIAGEAGFRVEVSSARQMSONGAL 2400
 2177 SIQ--KLSSIRKGLVHLGIFRADEGHLLIVLHLLVVDG-VSWRI-----L 2220
 2401 DAVFHHCCSO---GRTL-VNFPDTHLGRSDLLTNRPLQRLQNRRIAIEVRRLRSLPLS 2456
 2221 FEDFETLYSQALKGQTEIGYKTDYQEEFA-----RLKAYAHRSRTLSKAEAYWRNIKA 2275
 2457 ---YMTPSNIVLDMKPLNANGKVDKREKSRRAKVVPKQOOTAAPLPTFISEVEVILCEE 2513
 2276 RVRFIPKPNVL-----KEDVYENS-----TTLSIKLGE 2304
 2514 ATEVFGMKVDI---TDHFFNLGHSLLATKLISRI-----DQRLKVRIT---VKDVPD-- 2560
 2305 AT-----ADLLRNTNRAYNTEINDILLTALTGARDITGENKLKVMMEGHGEDTLEGV 2358
 2561 -----HPVFADLA--SVIROGL-----GLQOPVSDGOGDRSAHMAPRTEA 2601
 2359 DITRTIGWFTMYPVLLDAGEEAKLSQIKMWKETLRKIPNKGIGYGLLKMYA---EDPD 2415

Db 3654 GTGYD-----PPRNEIERKLVQVWREILGAEIDIGISHHFFAAGSDSKALQIVSRL-- 3704
QY 2198 GQVRSQATNEHFLAARAIHTLGNATKDDVROKMAELEDMEEBELL-VEPAFTTSLKDRFP 2256
Db 3705 AKNNLKEMKALFANKIDLSRFEITEETHRKHNPVTGETETELPIQKRYFANNKEELD 3764
QY 2257 GLVHEVILPKMEAVNEL-SAYRYAAVHVVRGSLGDELVLPEKDDWIDFQANLNQKS 2315
Db 3765 HFNQSFMLFRKGDYDENIVRTAFNKILEQH-----DALRMIYEEDGDGTIOYNGRYEN 3818
QY 2316 LGDLLKSSDAAINAVSKIPETITAFERQVVASLNSNIDWOLSTIRSSAEGDSSLSVPDI 2375
Db 3819 LFDL-----DV-----YDVRGFSQ-----EEKVPELATGQKSSSI----- 3950
QY 2376 FRIAGEAGRVEVSSAROWSQNGALDAVPHHCCSQGRTLVNFPDTHHLRGSLLTNRLQ 2435
Db 3851 -----RKGKLVHLGIFRADEGDHLLAI-HHLVVDG----- 3880
QY 2436 RLONRRIAEVRRLRLSLPSYMPISNIVLDKMPLNANGKVDK-KELSRRAKVPKQ-- 2493
Db 3881 --VSMRILFEDFETL-----YLAQKGEPLDIGYKTDYSQEFARQLKKAQSR 3927
QY 2494 -----TAAPLTPPISBEVILCE-EATEVFGMKVDITDHF-----FNLGGH 2534
Db 3928 LKEREYWKALEADVPFPAEKLERDTEHSATLSIRIGPDVDTAKLLNFAKAYNTEIN 3987
QY 2535 SLLATKLISRIDQRLKVRITKVDVDFHVPFADLASVIRQGLGQVPSDQGDGSAHMA 2594
Db 3988 DILLTALIA-----AVRDIITGE-----NKLKVMMEGHG-REDILDGV----- 4023
QY 2595 PRTEATAILCDEFAKVLGVQVGTIDNFFDLG-----GHSMLATKLAVER-----IGH--- 2640
Db 4024 -----DITRIGFTVTPYVFDLGEKEISQNIKMKVKEALRIPNKGIGYVL 4072
QY 2641 RLDTTVSVKDVDFHVPFLQALALDNLVQSKT-----NEIVGGR----- 2679
Db 4073 KYMTEELQKIQOTQAPLSFNFGEMNDNRKVFSPFSGESIGGKIYRHCALMENNAIS 4132
QY 2680 ---EMAEYSPF-----QL--LFTEDPEFM-----ASEIKQO-----LE 2708
Db 4133 LAGELTIYTFNODQYQTSITIEOLNGSFENLEKIVDHCVDKESDMPSTDYGDVSLGLE 4192
QY 2709 LQIIOO-----IYPTOMOKAFLE---DHTTARPRPVFPVYIDFPSPSEPDAG 2755
Db 4193 ELELIKDKYSAFQIEKIYPLANNKGMFLHNMADQISG--AYQOIVIKLGRVHDI-- 4248
QY 2756 LIKACESLHNLDFRTVFA-EASGELYQVLSCLDLPQIVIEDTINATNEFLDEFA 2814
Db 4249 LEESFHEIVKRHEILRASFEYEITAEPROIARDKRTPTSIDLTGENRTRQHRFETVL 4308
QY 2815 KEPVRLG-----HPLIRFTIIKOT-KSMRVIMRISHALVDGLSLHVRVKLHMLY----- 2863
Db 4309 KEDQERGFDSLSSPALMRVCLIKMSDESRYRLIWSHHILLDGCWGLGVLSFLSYLKIMK 4368
QY 2864 -NGRSLPLPHOFESKRYMOY-TADGRESGHGFWRDVION-----TPMTILSD 2907
Db 4369 GESRRLKEPKPYGDYIKWLEKQDQEBAYVWKDYLGKYSRSELPAFNKATSEECYCGE 4428
QY 2908 TVVD-GNDATCKALHLSKIWINIPSSQVLRGSSNIITQATVFNAAALVLSRESKDVWFG 2966
Db 4429 KVISFSKELTKITRIAK-----QHHVTINTVLQGIWGMILAKYKNTDEVFG 4476
QY 2967 RYVSGRQLPVE-YQDIVGCTNAPVRAHIES-DYNQLLHDIQOYLLSLPHEITIGFS 3024
Db 4477 TVVSGREA-PVQIEEMVGLFIHTIPTRISFEGARSFKVLLKTKQAESIESNYSYNNLS 4535
QY 3025 DLKRNCTDPEATNFSCCITYHNEFYHPESQEQVQVEMGLTKFVNTEMDEPLYDLAI 3084
Db 4536 EIOVLSMKRELTH-----VMAFONYAF-DEELFRSQSGETGFELEGVH-GKERTNTNPNL 4590
QY 3085 AGEVFPDGAGLK-----TVIAKTOLFGKRKRVEHLEEVSKTFEGNL 3126
Db 4591 TGVLEDEQLKLTFTENYVDNTIETLEKHIITVAEQVADETTQTLRDIN 4641

RESULT 4

TYCB_BACBR ID TYCB_BACBR STANDARD; PRT; 3587 AA.
AC G30408;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase II [Includes: ATP-dependent proline adenylase
DE (ProA) (proline activase); ATP-dependent phenylalanine adenylase
DE (PheA) (phenylalanine activase); ATP-dependent D-phenylalanine
DE adenylase (D-PheA) (D-phenylalanine activase); Phenylalanine racemase
DE [ATP-hydrolyzing] (EC 5.1.1.11)].
GN TYCB.
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 8185;
RX MEDLINE=98012987; PubMed=9352938;
RA Mootz H.D., Marahiel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
nucleotide sequence and biochemical characterization of functional
internal adenylation domains.";
RT J. Bacteriol. 179:6843-6850(1997).
RL J. Bacteriol. 179:6843-6850(1997).
CC -!- FUNCTION: ACTIVATES THE SECOND TO FOURTH AMINO ACIDS IN TYROCIDINE
CC (IN TYROCIDINE A, PRO, PHE, AND D-PHE) AND EPIMERIZES THE LAST
CC ONE.
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine = AMP + diphosphate + D-
CC phenylalanine.
CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES (BY
CC SIMILARITY).
CC -!- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC -!- SUBUNIT: LARGE MULTIZYME COMPLEX OF TYCA, TYCB AND TYCC.
CC -!- DOMAIN: CONSISTS OF THREE MODULES, INCLUDING AN C-TERMINAL
CC EPIMERIZATION DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
CC N METHYLATION (OPTIONAL).
CC -!- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF004835; AAC45929.1; -
CC HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00668; Condensation; 4.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE; PS00455; AMP-BINDING; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
KW Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.

| | | | | | |
|--|----------|---|---------------------------------------|--|--------|
| FT | REPEAT | 466 | 1045 | DOMAIN 1 (PROLINE-ACTIVATING). | |
| FT | REPEAT | 1522 | 2081 | DOMAIN 2 (PHENYLALANINE-ACTIVATING). | |
| FT | REPEAT | 2540 | 3122 | DOMAIN 3 (D-PHENYLALANINE-ACTIVATING). | |
| FT | DOMAIN | 977 | 1044 | ACYL CARRIER (ACP) 1. | |
| FT | DOMAIN | 2012 | 2079 | ACYL CARRIER (ACP) 2. | |
| FT | DOMAIN | 3045 | 3111 | ACYL CARRIER (ACP) 3. | |
| FT | BINDING | 1007 | 1007 | PHOSPHOPANTHETINE (BY SIMILARITY). | |
| FT | BINDING | 2042 | 2042 | PHOSPHOPANTHETINE (BY SIMILARITY). | |
| FT | BINDING | 3075 | 3075 | PHOSPHOPANTHETINE (BY SIMILARITY). | |
| SQ | SEQUENCE | 3587 | AA; 404810 | MW; 322B8471BBB28B47 | CRC64; |
| Query Match | | | | | |
| Best Local Similarity 11.3%; Score 1816; DB 1; Length 3587; | | | | | |
| Matches 844; Conservative 577; Mismatches 1359; Indels 1048; Gaps 140; | | | | | |
| QY | 31 | SYEQVFLYGLDSSRIEAIKPCPTPQLDMIDCNALDKQ-- | SAIGHAVYDVPTDIDISREA | 88 | |
| DB | 5 | SKEQVDMYAL----- | TPMQEGMLFHALLDOEHNSHLVQMSISLOGDLVGLFT | 53 | |
| QY | 89 | LAWKEIVNQTALPAFAFTSDSGKTSQVILKDSFV-FS | WMCSSSSSPDEVVRDEAAAA | 147 | |
| DB | 54 | DSLHLVRYDVFTLFLEYEKLKPLQVVLKQRPPIEFY | DLSACDESEKQLRYQYKRA | 113 | |
| QY | 148 | SGPRCNRF----- | VLLDMQTKKCOLWTFESHALVDVTFQORVLSRVFAAY | KHEKOTH | 200 |
| DB | 114 | DOERTFHAKDPLMRVALFQMSQHDYQVWSPHILMDG | WCFSIIFDDLLATLYLSQN-- | 171 | |
| QY | 201 | RPTPESSDATDTSQSVSVSMSCEDNAVSAFWQTHLN | DNLNASVPFPHLSDHLMVNP | 260 | |
| DB | 172 | --KTALSLEVPQYSRFINWLE--KONKQALNFW--- | SDYLEAYEQ | 211 | |
| QY | 261 | TTT----- | AHRITFP----- | LSQKALSNAICRTALSILSRVTH | 296 |
| DB | 212 | KTLTPKKEAFAKAFQPTQYRESLNRRTLKQLTIASQ | NOVTLSTVIQITWGLVLLQYNA | 271 | |
| QY | 297 | SDAELFGAVTE--- | QSLPDKHYLADGTYQTVAPLRVCHQCNLRASDVMDA | ISSYDDL | 352 |
| DB | 272 | AHDVFGSVSGRPTDIVGIDKMV--GLFINTIPRQAKA | GOTFSELLQAVHK----- | 323 | |
| QY | 353 | GHLAPFGLDIRNTDNGSAACDFQTV----- | LLYTDGSHVNNGIN----- | G | 394 |
| DB | 324 | ----- | RTLQSYEHPVLDIQTQSVLQKQELIDHLVNIENYPL | VEALQKALNQOIG | 375 |
| QY | 395 | FLOQITESHFPCNNRALLLHCQMESSGALLVAYYD | HNVNVDLSQTLRLLOQF-GHLIKC | 453 | |
| DB | 376 | F--TITAVEMFEPTNYDLTVVMPKEE--- | LAFRFDYNA--ALFDEQVQVQKLQAGLQOI | 427 | |
| QY | 454 | LQSPDLDS--SMAEVLNLTEDYDRAETESWNSQPLE | VQDPLIHHEML-KAVSHSPTKTAIQ | 510 | |
| DB | 428 | ADCVANNSGVELCQIPLLTAEATSQLAKRTETAADY | PAATWHELFSSQAKTPEQVAVV | 487 | |
| QY | 511 | AWDGDWYSELNVSRLAVHIKSLGLKAQQAIIIPVY | FEKSKWVIASMLAVLKSNAFTL | 570 | |
| DB | 488 | FADQHLTYRELDEKSNQALARELRKGI-CTGSLVGT | LLDRSLDMIVGLVKGAGAFVP | 546 | |
| QY | 571 | IDPNPPARTAQVTTQATRVALTSLKHRETV---- | OKLVGRCVVVDDELLOSVSASDDF | 626 | |
| DB | 547 | IDPELPAERIAIYMLTHSRVPLVVTQNHRAKVTPTT | ETIDINTAVIGEE----- | SRAP1 | 600 |
| QY | 627 | SSLTKSODLAVYFTSGSGDPKGMIEHRAFSSCALKF | GAISGINSDDTRALQFTHAFG | 686 | |
| DB | 601 | ESLNQPHDLFYIYISGTTGQPKGVNLEHNRNANLHM | HFTFDQTNIAFHEKVIQYTCSPD | 660 | |
| QY | 687 | ACCLEIWTTLINGCVCIPDDDR--MNSIPSFNRYNNV | MMATPSYMG-TFSPEDVPG | 743 | |
| DB | 661 | VCYQETFSLLSGQLYLITNELRRHVEKLFQAEKQIS | ILSLPVSFLKFTFNEQDYAQ | 720 | |
| QY | 744 | -----LATLVLVGEQMSVNAIMAPKIQ----- | LLNGYGOSESSSICFASNMSTE | 789 | |
| DB | 721 | SPPRCVKHIITAGEQL----- | VVTHELQKYLQHRVFLNHYGSPETHV-- | TTCTMD | 771 |
| QY | 790 | PNN----- | MGRV-GAHSHWIDPNDINRLVPIGAVGELVIESPG | IARDYIVPPPPKS | 841 |

| | | | | | |
|----|------|---|---|----------------------------------|------|
| Db | 772 | PQQAIPPELPPICKPISNTGIYILDEG-- | LQKPEGIVGELYISGANVGRGYLHQP | 829 | |
| Qy | 842 | PFETDIPSWYANTFPDGAKLYRTGDLARYASDGSIVCLGRIDSDQVKIRGORVELGATET | 901 | | |
| Db | 830 | KFLD----- | NPYQGERMYRTGDLALWLPDQGLEFLGRIDHQQVKIRGRHIEUGETES | 881 | |
| Qy | 902 | HLRQMPDOLLTIVWEATKRSQNSASTSLIAFLIGSSYFGNRPDSADHILHDHDAKAINIKL | 961 | | |
| Db | 882 | RLLNHPAIEKAAVVID-- | RADETGGKFLCAYVVLQKALSDEEMRAY----- | L | 925 |
| Qy | 962 | EQVLPKHSIPSYICMLPRTATGKIDRRRLRMCKDKILDQKQGAIVQOAPAPIPVFA | 1021 | | |
| Db | 926 | AQALPEYMPISFFVTLERIPVTPNGKTRRAL----- | PKPEGSAKTRADYVAPT-T | 975 | |
| Qy | 1022 | DTAAKLHSIWQSLGIDRATVNGATFFELGNSITAIKMYN-WARSVGMGLKVSNIYQH | 1080 | | |
| Db | 976 | ELEOKLVAIWEIILGVSP-- | IGIODHFTFLGGHSLKAIOLISRIQKECOADVPLRVLFQ | 1033 | |
| Qy | 1081 | PTLAGISAVVK-GDPLSYLTPKSTHEGVPESQSGRLWFLDQLDVGSLWLYLIPYAVRM | 1139 | | |
| Db | 1034 | PTIQALAAVVEGESAYLAIPAEPQAYYPVSSAQKRMILLNQLDPHSTVYNLPMVAIL | 1093 | | |
| Qy | 1140 | RGPNVNDALRALAALAEQRHETLRTTFEDQDGVGVQIVHEK----- | LSEEMKVIDLC | 1191 | |
| Db | 1094 | EGTLDKARLEHAISNLVARHESLRTSFHTINGEPVSRHIEQGHLPVILETAEOQVNEVI | 1153 | | |
| Qy | 1192 | GSDLPDFEVLNOBOTTFFNLSEAGWRATLLRLGEDDHILITVMHHIISDGWSIDLVRD | 1251 | | |
| Db | 1154 | LGFMQPFOLV--- | TAPL----- | CRGLVLAENRHLVITDMHHIISDGVSQILIND | 1202 |
| Qy | 1252 | LNQLYSAALKDKOPLSALTPLIOYSDFAKWK-- | DQFIEQEKQLNWKQKLDSSPA- | 1308 | |
| Db | 1203 | FSRLYQN----- | KALPEORITHYKDFAWWEKAWTQTDYQKQKYLWDRFAGEIPVL | 1253 | |
| Qy | 1309 | KIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFCEHNTTTFVLLAFAAHRYLRTAV | 1368 | | |
| Db | 1254 | NLPMDYPRPAVQSFEGERYLFRTKOLLESLODVAQKTGTTLYMWLLAAHYHLLSSYSGQ | 1313 | | |
| Qy | 1369 | EDAVTGTPIANRNPEDIIIGCFVNTQCMRINIDHHDFTGTLINQVWATTTAAFNEDI | 1428 | | |
| Db | 1314 | DOVMIGTVTAGRVHPTESMTGMFVNTLAMRQSAPTKTRQFLLEVKADNTLAAFEHQY | 1373 | | |
| Qy | 1429 | PFERYVALQPSRDLSSPTLAQLIPAVHSQKDLGRFKQGLSEVP-VPSKAYTRFDMF | 1487 | | |
| Db | 1374 | PFEELVEKLA1-QRNRNPLFDTLFILONM-DADLIELDGLTVTYPYVEGEVAKFDLSL | 1431 | | |
| Qy | 1488 | HLFOETDLSKGSVNFADLEFKMETVENVVFFFEILRNGLQSSRTPSILPDTGIVTLE | 1547 | | |
| Db | 1432 | EASENQAGLSFCFECKTLFARETIHERMSLHLYLILO-- | AVSANTQEOLAQI-EMLTAAHE | 1488 | |
| Qy | 1548 | KLDVL-- | NVKHVDYPPRESSLADVFQTVSAYPDSLAVVDSSCRUTYTEDRQSDILAGW | 1604 | |
| Db | 1489 | KQELLVHFNDAALYPAESTLSOLFEDQAKTPEQTAVVFGDKRLTYRELNERANQLAHT | 1548 | | |
| Qy | 1605 | LRRRMPAETLVAVAPRSCETIVAFGVKLANLAYPLDVRSPSARVQDILISGLSGPTI | 1664 | | |
| Db | 1549 | LRAKGVQAEQSGVIGMAQSRLEMAIGI IAILKAGGAYVPIDPYPNERIAYMLEDCR-- | RL | 1606 | |
| Qy | 1665 | VLIGHDTAPPDLEVTNVEFVIRDALNDSNADGFEVIEHDSKTP-- | SATSLAYLVYTSGS | 1722 | |
| Db | 1607 | VLTQOOLA-- | EKMANTVECLYL----- | DEEGSYSPQTEPIETPIHTAADLAYIYTSGT | 1657 |
| Qy | 1723 | TGRPKGMIEHRYIIRTVT----- | SGCIPNYPSETRMAHMAITAFDGSASYEY | 1770 | |
| Db | 1658 | TGRPKGMVHEHRYIIRTVT----- | SGCIPNYPSETRMAHMAITAFDGSASYEY | 1770 | |
| Qy | 1771 | SALLFGRTLVCVDYMTTLARALKOVFFREHVNAAHSHVTSQDQVPLVRPRLSRTLMFF | 1830 | | |
| Db | 1707 | TLIVSGSTVVLMPDHEAKDPIALRNLIAWECYVVFVPS----- | MF | 1748 | |
| Qy | 1831 | FLVVTSTAPDALDAQGLYQG----- | VQCYNGYGTENGVMSTIYPI | 1872 | |
| Db | 1749 | QAILCECSTPADIRSTQAVMLGGEKLSPKLVQLCKAMHPQMSVMNAYGPTESSVMAT-YLR | 1807 | | |

1873 DSTESFNGVPIGRALNSGAYVVDPEQQLVGIGVMGELVVTGDLARGYSDK-ALDENR 1931
1808 DTQPD--OPTITIGRPIANTAIYVDQHHQLLPVGVGEICIGGHLARGYWKPELTAEK 1865
1932 FVHTVNDQT--VKAYTGTDRVYRIGDGLIEFGMDTQFKRGNRIESAETEAALLRD 1989
1866 FV---ANPVPGERMYKTGDLGRW-LHDGTIDFIRVDDQIKVGRYRIEVEGEIEAVILL 1921
1990 SSVDAAVLQONE----- 2003
1922 DQINEAIVVAYQDGRGDSYLAAYVTGKTAIESELSRAHLRLRELPAYMVPYLIQLDAFPL 1981
2004 -----DOAPEILGFVVADH----- 2017
1982 TPNGKVDKALPKPEGPATGAAYVAPATEVEAKLVAIWENALGICSGVGLDHFELGGH 2041
2018 -----DHSENDKQS----- 2027
2042 SLKAMTVVAVQHREFQIDLLKQFFAAPTIRDLARLIEHSEQAAGAAIQPAEPQAYPVS 2101
2028 -----ANQVEGWQDHSES-GMYSDIGEIDPSTIGSDFKGTSMYDGSQIDFDEMHEW 2078
2102 SAQORWYLLHQLLEGAGISYNTPGIIMLEGKLDREQLANALQALVDRHIDILRTSEWVGD- 2160
2079 LGETTRTLHNRSLGNVLEIGTSGMILFNLDLSRLESYVGLPSR-----SAAAFVNKATES 2135
2161 --ELVQKIHDR-----YAVNME-----YVTAEEQQIDDLFHFVVRPFDLS 2198
2136 IPSLAGKAKYQVG-----TATDIGOVDDLHPLAVL-----NSVIOY----- 2172
2199 VPPELLRMSVLKADERHLLYDMHHAADAASITILFDELAELYQGRELPDEMRIQKQDFA 2258
2173 -----FPSSE--YLAETADTLIHL-----PNVQRIFFGDVRSQATNE----- 2207
2259 VMQKALHESDAFKQEAAYWLSTFAGNITAVDPTDPRPAVKFAGQVTLSDMQELLSA 2318
2208 -HFLAARAIHT-----LGNKATKDDV-----RQMAE----- 2233
2319 LHLEAAHTNTLFWLLAAYNVLLAKYAGQDDIIVGTPISGRSKRAELAPVGVMEVHTLAI 2378
2234 -----LEDWEEELLVEPAF-----FTSLKD-----RPPG-----LVE 2260
2379 RNKPTAEKTFQLEQVKQNAL--DAFDHQDYFESLVEKLGIPDPRGNPLFDFTMFIQ 2436
2261 HVEILPKNM-----EAVNELSAYRYAAVHVHVRGSLGD----- 2292
2437 NDELHAKTLQLVYRPVESDSALDVAKFDLSFHLTERETDLFLURLEYCTKLFKQQTVERM 2496
2293 -----ELVLPVEK-----DDWIDFQANLQNO-----KSLGDL 2319
2497 AHFLQILRAVTANPENELQEIEMLTAAEQMILLVAFNPDTHREYRADQTIQQLFEELAEK 2556
2320 LKSSDAAIMAVSKIPF-----EITAFER-----QVVASLNSNIDEMOLSTIRSSA 2364
2557 MPHTALVPEEKMSFRELNERANQALAVLRKGVGPAQIVALLVRSABEWITATLTK 2616
2365 EGDSSLSV-PDI--FRIAGEAGFRVSVSSARQWSQNGALDAVPHCCSQGRTL----- 2414
2617 AGGAFLPVDPDYPEERIR-----YMLEDSQAKLVVTHAHL---LHKVSSQSEVVVDVDDPGS 2669
2415 -----VNPTD-----HHL----- 2423
2670 YATQTNLPCANTPSDLAYIYVTSGTGPKGVMLHKGVANLQAVFAHHLVTPQDRAG 2729
2424 -----RGSDDLNTNRLQRLQN-----RRIAEVRELRSL---PSYMI-- 2459
2730 HFASISFDSAVMDMFGPLLSGATLYVLSRDVINDFORFAEYVRDANITELTLPPTAIYL 2789
2460 -----PSNIVLDKMP-----LNANGKYDKRSLSRRAKVVYPKQQTAAPL 2498
2790 EPEQVPSRLITAGASSVALVDKWEKVTYVNGYGPTESTVCATLWAKAPDE----PV 2845

2499 PTF-----PISEVEVILCEEATEV-----FGMKVDIITDHF-----NL 2531
2846 ETITIGRPIQNTKLIYVDDQLQALQPMQMGELCISGLARGYWNRELTAKEFVDNPFV 2905
2532 GGHSLAT-----KLISRIDQRLKVR-----ITVKDVF----- 2559
2906 PGTMYRTGDLARWLDPDGTIEYLGRIDHQVKIRHVRVELGEVSVLLRYDTVKEAAALITH 2965
2560 --DHPVFADLASVIRQGL-----QOPVSDGQGOGRSA- 2591
2966 EDRGQAYLCAYVAEGEATPAQLRAYMENELPNMVPAPAFIQLEKMLPLPNDKIDKAL 3025
2592 -----HMAPTETEAILCDEFKVLGF-QVGITDNEFFDLGGHSLMATKLAURI 2638
3026 PKPQNEENRTEQYAAQPOTELEQLLGIWADVGLIKQVGTQDNFFELGDSIKAIQVSTRL 3085
2639 GHRLDITVSVKVDFHPVLQALALONLVOSKTNEIVGGREMAEYSPFQLLFTEDPEEF 2698
3086 -NAGWMTLAMELFOYPTIEEAALV-----IPNSRE----- 3116
2699 MASEIKPQLELQELIODIYPTOMOKAFLFDHTTAR---PRPFVFPYID-FPSTSEPDAA 2754
3117 --SE-----QGVVEGEIALTPIQKWFANFTDRHHNQAVMLFREDGFDE----- 3160
2755 GLIK-ACESLVNHLDIRTVFAEASGELYQVWLSCLDLPQIVTETEDNINTATNEF----- 2809
3161 GLVROAFQOQIVEHHDALRMVYKQEDGAIKQINRGLTDERFR-FYSYDLKNHANSEARILE 3219
2810 LDEFKPEVPLGH-PLIRFTIITKQSMRVMIRSHALYDGLSLEHVVRKL-----HMLY 2863
3220 LSQIOSSIDLEHGLHVHVALFATKGDHLLVALHHLVVDGVSWMRILFEFESSAYSQALH 3279
2864 NGRSLAPP-----HOFESRYMQYTADGRE--SGHGFWRDVIQNTPTMTILSDTVVDGNDAT 2916
3280 QQEIVLPKTKDSFKDWAQAQKYVADSDELLREYAVWHNLETTTITTAALPTDFVT----AD 3335
2917 CKALH---LSKIVNIPESQVLRGSSNIITQA-----TVFN-----AACALVLSRESDSKDV 2964
3336 RKOKHTRTLSFALTVPQ-----TENLLRHVHHAYHTEMNDLLLTALGLAVKDWHTNGVV 3390
2965 FGRIVSGROCLPVEYQ-DIVGCTNAVPVRAHIESDYQNLHDIOD--OYLLSLPHET 3020
3391 INLEHGREDIQENMVNTRTIGWFTSQYPVVLDMEKAE--DLPYQIKOTENLRRIKPKG 3448
3021 IGFSDLKRNCTDW---PEAITNFSCCITYHNFYHPSESQEQOQVEMGVLTKEVNIEMDE 3077
3449 IGVEILRTLTSQQLPPLA-----FTLRPEISFN----- 3477
3078 PLYDLAIAGEVEPDG--AGLKVTVIATQTLFGRKRVHLLLEEVSKTFE 3123
3478 -----YLGQFESDGTGGTFTSPLGTGQLFSPESEVFLDISAMIE 3519

RESULT 5

PPS3_BACSU ID PPS3_BACSU STANDARD; PRG: 2555 AA.
AC P39847;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptide synthesis 3.
GN PPS3 OR PPS3.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RL Mol. Microbiol. 8:821-831(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=168;
 RC MEDLINE=97124189; PubMed=8969502;
 RA Yamane K., Kumano M., Kurita K.,
 RT "The 25 degrees-36 degrees region of the Bacillus subtilis
 chromosome: determination of the sequence of a 146 kb segment and
 identification of 113 genes."
 RL Microbiology 142:3047-3056(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogawara N., Mosser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
 RA Entian K.D., Errington J., Fabret C., Ferrali E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinou S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scantlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis."
 RL Nature 390:249-256(1997).
 RN [5]
 RP SEQUENCE OF 514-800 FROM N.A.
 RC STRAIN=ATCC 21332;
 RX MEDLINE=92290255; PubMed=1601288;
 RA Borchert S., Fatil S.S., Marahiel M.A.;
 RT "Identification of putative multifunctional peptide synthetase genes
 using highly conserved oligonucleotide sequences derived from known
 synthetases."
 RL FEMS Microbiol. Lett. 71:175-180(1992).
 CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
 CC
 CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
 CC
 CC -1- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
 CC
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY
 CC
 CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: D13262; BAA02523.1; -;
 CC EMBL: X70356; CAA49817.1; -;
 CC

DR EMBL: D50453; BAA09883.1; -;
 DR EMBL: Z99105; CAB12143.1; -;
 DR EMBL: X65835; CAA46678.1; -;
 DR HSP: P14687; LAMU.
 DR Subtilist; BG10169; srfAB.
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR001242; Condensatn.
 DR InterPro: IPR003880; Ppantne_attach.
 DR Pfam: PF00501; AMP-binding; 3.
 DR Pfam: PF00550; pp-binding; 3.
 DR Pfam: PF00668; Condensation; 4.
 DR PRINTS: PRO0154; AMPBINDING.
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 3.
 DR PROSITE: PS00455; AMP BINDING; 3.
 DR PROSITE: PS0075; ACP_DOMAIN; 3.
 KW Ligase: Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
 KW Multifunctional enzyme; Repeat; Complete proteome.
 FT REPEAT ? 1040 DOMAIN 1 (VAL-ACTIVATING).
 FT REPEAT ? 2096 DOMAIN 2 (ASP-ACTIVATING).
 FT REPEAT ? 3114 DOMAIN 3 (D-LEU-ACTIVATING).
 FT DOMAIN 1036 ACYL CARRIER (ACP) 1.
 FT DOMAIN 2015 2082 ACYL CARRIER (ACP) 2.
 FT DOMAIN 3043 3109 ACYL CARRIER (ACP) 3.
 FT BINDING 999 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT BINDING 2045 2045 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT BINDING 3073 3073 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT CONFLICT 33 33 S -> F (IN REF. 1).
 FT CONFLICT 42 42 A -> G (IN REF. 1).
 FT CONFLICT 110 110 Q -> D (IN REF. 1).
 FT CONFLICT 113 115 ROA -> ACG (IN REF. 1).
 FT CONFLICT 139 139 A -> V (IN REF. 1).
 FT CONFLICT 259 259 L -> W (IN REF. 1).
 FT CONFLICT 309 309 R -> A (IN REF. 1).
 FT CONFLICT 478 480 TPA -> SRP (IN REF. 1).
 FT CONFLICT 596 596 MISSING (IN REF. 5).
 FT CONFLICT 648 648 A -> R (IN REF. 1).
 FT CONFLICT 680 682 RHV -> ETL (IN REF. 1).
 FT CONFLICT 694 698 EOSIT -> DKRIS (IN REF. 5).
 FT CONFLICT 788 788 M -> L (IN REF. 5).
 FT CONFLICT 939 940 PL -> LV (IN REF. 1).
 FT CONFLICT 1038 1038 N -> I (IN REF. 1).
 FT CONFLICT 1133 1133 H -> Q (IN REF. 1).
 FT CONFLICT 1310 1310 V -> C (IN REF. 1).
 FT CONFLICT 1333 1333 G -> V (IN REF. 1).
 FT CONFLICT 1384 1384 P -> R (IN REF. 1).
 FT CONFLICT 1582 1582 G -> E (IN REF. 1).
 FT CONFLICT 1677 1682 KRRADG -> E (IN REF. 1).
 FT CONFLICT 1700 1700 C -> S (IN REF. 1).
 FT CONFLICT 1755 1755 F -> K (IN REF. 1).
 FT CONFLICT 1787 1787 T -> S (IN REF. 1).
 FT CONFLICT 1801 1822 GAIAGRVDLYEPDAFAKRPITG -> APSPGGLICMSRCIC
 FT ETPDNR (IN REF. 1).
 FT LG -> PK (IN REF. 1).
 FT R -> C (IN REF. 1).
 FT A -> V (IN REF. 1).
 FT ARLTP -> LRSLN (IN REF. 1).
 FT E -> Q (IN REF. 1).
 FT ATDLF -> RQCS (IN REF. 1).
 FT TVHOLFETVVRHKRPVTV -> DGCISYKRLSSATKTK
 FT ARLSHT (IN REF. 1).
 FT MSAALVGV -> KCPRCAS (IN REF. 1).
 FT KL -> NV (IN REF. 1).
 FT D -> H (IN REF. 1).
 FT D -> H (IN REF. 1).
 FT GELCVA -> RALRG (IN REF. 1).
 FT RF -> L (IN REF. 1).
 FT EDK -> QDA (IN REF. 1).
 FT R -> A (IN REF. 1).
 SQ SEQUENCE 3587 AA; 401248 MW; 6B0B05A9FF32054D CRC64;
 11.1%; Score 1793.5; DB 1; Length 3587;
 Best Local Similarity 22.4%; Pred. No. 2.8e-95;
 Matches 732; Conservative 458; Mismatches 1130; Indels 953; Gaps 99;

QY 41 LSSRIEAIKCTPQFQDMIDCNALDKOSA-----IG-HAVYDVPTDIDISFALAWK 92
D 1 MSKKSIOKVYALTQMEGLMYHMLDPHSSSYSTQLELGHIAAF-----DLEIFKSVN 54
QY 93 EIVNOTPALRAFAFTSDSGKTSVILKDSFVSNMSSSSSPDE-VVROEAAAASGPR 151
D 55 ELIRSYDILRTVVFHQLOKQPRVYLAER--KTVHYEDISHADENKQEHIERKYKOVQ 112
QY 152 CNREVLEDMQTK-----KCOLVMTFESHALVDVTFQORVLSRVFAAYKHEKDHTRP 202
D 113 ROAFNLAKDILFKVAVFLAADOLYLAWSNHHIMMDGSMGLMKSLSFQNYEALRAGRT 172
QY 203 ETPESSDATDSDOSVWSMCDNAVSAATHFQWTHLNDL-NASVFP-----249
D 173 ANGCKPYSD-----YIKLWGDONE-EAESYWSERLAGFEQPSVLPGRLPVYKDEYVN 225
QY 250 ----HLSHMLMNPPTTAE-HRTFFPLSQALSNSAICRTALSILLSRYTHSDLEALFGA 304
D 226 KEYSFTWDELVARIQOTANLHQTGP-----NLFQAVLGIVLSKYNFTDDVIFGT 276
QY 305 VTEOSLPFDKHYLAD--GTQVTAPLRVHCQSNLRASDVMDAISS-----YD-----349
D 277 VV-SGRPSEINGIETMAGLFNTIPVRVKVERDRAFIETAVQOHAVERDYVPLYE 335
QY 350 -----DRLGHLAPGLRDINTGNGSACDFQVLLVYTDGSHVNGINGFLQOIT 400
D 336 IQKRSALDGNLLHLVAFENYPLDQLENGSMEDRL-----GF--SIK 376
QY 401 ESSHFPCNNRALLHCOMESSGALLVAY-----YDHNVIDSLQTRLLQOFGHL 450
D 377 VESAF-----EQTSFDFNLIVPCKTWVKIKYNGAFAFSAFTAE---HL 421
QY 451 IKLOQSPD--LSSMAEVNLM-TEYDRAETESNNSQPLEVODTLIHML-KAVSHSPK 506
D 422 TRMEEAVIDQPAFVREYGLVGEQRIIVEFENSTKAELPEGMVHOFVEEQAKRTPAS 481
QY 507 TAIQAWDGMTYSLELDNVSRLAVHIKSLGRQAQALIPYVFEKSKWVIASMLAVLKSNG 566
D 482 TAVYETKLTLYRELNAANRLARKLVEHGLQGET-AAIMNDRSVETVVGMLAVLAKAGA 540
QY 567 AFLIDNPDPARTAQVVTOTRATVAL-----TSKLHRETQVLVGRVGVVDDELQSVS 621
D 541 AYVPLDPAALRFRMAEDSSVRVWLIGNSYTGQAHLQVP-----VLTLDIGFESE 594
QY 622 ASDDFSLSQDLAYVIFTSQDGPKGIMIEHRAFSSCALKFGASLGINSSTRALQFG 681
D 595 AADNLNLPAPSOLAYIMYTSQSTGPKGVMIHKSILR-LVKNAGYVVPVTEEDAMAQTG 653
QY 682 THAFGACLLBIMTTLINGGCVCIPSDDDRM---NSIPSFINRYNVNMMAATPSYMGTFSP 738
D 654 AVSFDACTFEVFGALLNGAAL-YPVKKRHVLDKQFAFLREQSITTMWLTSLPLNQLAA 712
QY 739 EDVPGIATL-----VLVEQMSVSNIAWAPKLOLLNGYQSSSSSCIFASNMSSTEP 790
D 713 KDAGMFGTLRHLIIGDALPHVIVSVKQA--SPSLSLWNGYGTENTFTSTFSLDREY 770
QY 791 NN--MGRVAG-AHSXVIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPEKSPFFTD 846
D 771 GGSIPKIGKIGNSTAIMDQOC--LOPIGAPGELCVGGVARGVYVNPPELFEKOFLED 838
QY 847 IPSWYFANTPPDGAKLYRTGDLARYASDGSIVCLGRIDSQVKIRGQVRELGAETHLROQ 906
D 829 -----PRPGERIYRTGDLARWLPDGNIEFLGRIDNQVKYRGFRIELGELETKL--N 878
QY 907 MPDDL--IIVVEATKRSQASNTSLAFILGSSYFGNRPNSDAHILHDHAKAINKLEQ 963
D 879 MAEHVTEAAVII---RKNKADENEICA-----YF---TADREV---AVSELKRTLSQ 921
QY 964 VLRPHSIPSPYICMLPRTATCKIDRRRLRINGKOILDQTOGAIVQOAPAPIPVFA-- 1021
D 922 SLPDYVWPAHLIQMDSLPTPNKINKKELP-----APOSEAVQPEYAAP 966

QY 1022 --DFAAKLHSUWOSLGIDPATVNVGAT--FFELGGSNITAIRM-VNMARSVGMDLKYSN 1076
D 967 KTESEKKLAEIWEIGILG-----VRAGVTDNFMIGGHSLKAMMTAKIOEHFHEVPKV 1021
QY 1077 IYQHTPLAGISAVVK-----GDPLSYTLIPKSTHEGPVEQSYSGRLFLDOLDVGLSWYL 1132
D 1022 LFEKPTIOELALYLEENESKEEQTFEPIRQASYOOHYVPSPAORMYILNOLGOANTSYN 1081
QY 1133 IPYAVMRGPVNDALRALAALRQTHETLRTTFEDODGQVQIVHEKLSSEMVIDLGG 1192
D 1082 VPVALLLEGEVDKDRLENAQOOLLNRHEILLTSEDMIDGEVQVIVHKNISHLEAAK--G 1139
QY 1193 SDDLPFVLENOEQTPPNLSEAGWRATLRLGLDEDDHILITVMHHIISDGMSIDVLRDL 1252
D 1140 REEDAEIIL-KAFVQPPFELNAPLVRSLKVLQLEKRHLLLIDMHIIITDGSSTGILIGDL 1198
QY 1253 NOLYSALKOSKODLSALTPLPIQYSDFAKWQKQDFIEQKQLNWKOLKDSFP-AKIP 1311
D 1199 AKIYOGA--DLELP-----QIHKYDYAVMHKEQ-TNYQKDEEYWLDFKGLPILDL 1248
QY 1312 TDFARPALLSGDAGCVHVTIDGELYQSLRAFNCBHNTTSFVLLAAFAAAHYRLTAVEDA 1371
D 1249 ADFERPAERSFAGERVFMGLDKQITAOIKSLMAETDTMYMFLAAFNLLSKYASODDI 1308
QY 1372 VIGTPIANRNPPELIEDIIGCFVNTQCMRINIDHDHDTGLTINQVKATTTAAFENEDIPFE 1431
D 1309 IVSPTAGRTHPDLQGVPMFVNTGALTAPAGDKTFAQLFEEVKTASLOAFERHQSYPLE 1368
QY 1432 RVYSALQPSRDLSSSTPLAOLIFAVHSOK-----DLGRFKFOGLESVPVPSKAYTRFD 1484
D 1369 ELIEKL-PLTRDTSRSLFSVMFNQMNWEIPSLRLGLDKIKISSYMLHHV-----AKFD 1420
QY 1485 MEHFLFOETOSLKGSNVFADELKMETVENVVRVFFELRNGLOSSRTPVSILPLTDGIV 1544
D 1421 LSLAEVREEDIGLSFDYATALEKDETIIRSRHFVNIK---AAAAANPVRLSDVDLLS 1477
QY 1545 TLEKLDVNLKVDYIPRESSLAOVFOQVSAYPDSLAVVDSRSSCLTTELDQRSQDIAGW 1604
D 1478 SAETAALLEERHMTQITEATFAALFEKQAOOTPDHSAVAGGNNLLTYRELDQANQLAHH 1537
QY 1605 LRRSRPAETLVAVFAPRSCETIVAFPGVLKANLAYLPDVRSPSARVQDILSGLSPTI 1664
D 1538 LRAQGAGNEDIVAIWDRSAEVVSVILGVKAGAAFLPIDPDPGERIRYSLED-SGAKF 1596
QY 1665 VLGHDTAPDPIEVTVNVEFVRIIDALNDSNAG---FEVIEHDSKPSATSLAYVLVTS 1721
D 1597 AVYNE-----RNMTAICQYEGIIIVSDDGKWRNESKERPSISSGRNLAIVYVTS 1647
QY 1722 STGRPKGMIEHRVLIIRTVTSGCIPNYP-----ETRMAMHATIAFDGAS 1766
D 1648 TTCKPKGVQIEHR-----NLTVVSWFSEAGLTKRRAGNDKTVLLSSVAFDLGY 1698
QY 1767 YEIYSALLFQRTLVCV--DYMTTLDARA-----LKDVFFREHVNAAASHVTS 1812
D 1699 TCMFPVLLGGELHIVOKETYTAPDEIAHYIKHEGITYIKLTPSLPHTIVNTASFADAN 1758
QY 1813 QDVPLRVPRLSRLTMFFFLVVTDSAPDALDAQGLYQGVQCYNGCTENGVMSTIPI 1872
D 1759 FE-----SLRLVILGGEKIIPTDVIAFRKMYGHTEFINHYGPTTATGAIAGRV 1807
QY 1873 D--STESFINGVPIGRNALNSGAYVVDPEOOLVGIGVMGELVVTGDLARGYSBK-ALDE 1929
D 1808 DLYEPDAFAKRPPIGRPIANAGALVLEAKLVPPGASGOLYITGGLARGYLNRPLTA 1867
QY 1930 NRVHITVNDQTVKAYRTGDRVRYRIGDGLIEFFGRMDTKFIRGNRIESAEIEAALLRD 1989
D 1868 EREVENYSPGSL-MYKTDGVR-RLSDGTAFIAGRADDQVKIRGYRIELGETETVMSL 1925
QY 1990 SSVYDAVYVLOONEDQAPILGFWADHDHSDKGSANQVQWGVQWDFESGMYSDIGE 2049
D 1926 SGIOE-AVVLAVSEGGQLQELCAYYTSDD-----1953
QY 2050 DPSTIGSDFKGMTSMYDGSQIDFDEMHEMELGTRTLHDNRSLGNVLEIGTSGMILFNL 2109

Db 1787 NMIGTETVHVTFKLSAADIANKSNIGRPLSTLQAHVMDAHNNLQFTGVPGLYICG 1846
 QY 1916 DGLARGYSDK-ALDENRNVH--ITVNDQTVAYRTGDRVRYRIGDGLIEFFGRMDTQFKI 1972
 Db 1847 EGVARGYLNRBELTADRVSNPLPGD---RLYRTGDLAK-RLSNGELEYLGRIDEQVKV 1902
 QY 1973 RGNRIESAEIEAALLRDSVRAAUVLQONEDQAPILGFVVADHDHSENDKQGSANQVE 2032
 Db 1903 RGHRIELGEIQAALLQYPMKEAAVITRADE-----1933
 QY 2033 GWQHFESGMTSDIGEIDPSTIGSKWTSWYDGSQIDFDEMHEWLGETTRTLHDNRSL 2092
 Db 1934 -----OGQTAIY-----1940
 QY 2093 GNVLEIGTSGMILFNLDLSRLSYVGLPESRSAAAFVNKATESIPSLAGKAKVQVGTATD 2152
 Db 1941 -----1940
 QY 2153 IGOVDLHPDLVNLVSVIYQFSPSEYLAIEADIHLPLNVQRIFFGDRVSQATNEHFLAA 2212
 Db 1941 -----1940
 QY 2213 RAIHTLKNATKDDVRQKMAELEDMEELLVEPAPFTSLKDRFPGCLVEHVEILPKNMEAV 2272
 Db 1941 -----1940
 QY 2273 NELSAYRYAAVHVHVRGSLDELVPVKDDWIDFOANQLNOKSLGDLKSSDAATMAYSK 2332
 Db 1941 ----AYM-----VIKDOQA-----1951
 QY 2333 IPFEITAFERQVVASLNSNIDEWOLSTIRSSAEGDSSLVDPDIFRIAGEAGFRVVSAR 2392
 Db 1952 ----NIS-----1954
 QY 2393 QWSQNGALDAVFHCCSGRGLVNPPTDHLHRLGSLDNLNRPLQRLQNRRIAEVRELR 2452
 Db 1955 -----DIRTYLKN 1962
 QY 2453 LLPSYMIPSNIVLDKMLNANGVDRKELSRRAKVPKQOATAPLTPPISEVEVILCE 2512
 Db 1963 ALPDMFLPMIQUIDSIPTVNGKLDQKALPE-----PEKA-----1999
 QY 2513 EATEVFGMKVDITDHFNLGGHSLATKLISRIDQRLKVRITVKDVPDHPVFIADLASVIR 2572
 Db 2000 -----VTADDI-----2005
 QY 2573 QGLGLQPVSDGQDRSAHMAPRTETAILCDEFKVLGF-QVGITONFDFGLGSHLMA 2631
 Db 2006 -----SPRNEIETVMAEIEELLNVDLGVSAFNFKLGDGSIKA 2044
 QY 2632 TKLAVRIGHR-LDTTVSVKDFDHPVLPQALADNLVQSKTNEIVGGRMAEYSPFOLL 2690
 Db 2045 LOVCARLQKRGFTT--VREMEHQTGLSARVKDVRADDOGPVEGE--ITWTPIQOW 2100
 QY 2691 FTEDPEEFMASEIKPQLELQEIQIDYIPSTOMQKALFDHTTARPPVPFVDPPTSE 2750
 Db 2101 F-----FSOSLESHFNQ-----SVMIYRAERFDE-----2125
 QY 2751 PDAAGLIKACBSLVNHLIDFRVFAEASGELYQVVLSCLDL-----PIQVETEDNINT 2804
 Db 2126 ---AALRVKLSLVTHDHALRVCHRDGGRQVQINRG-IDLSDEELYALELFDVDRKSLTE 2181
 QY 2805 ATNEFLDEFA---KEPVR-L-GHPLIRFTIIRKTSMRVIMRISHALYDGLSEHVVRKLH 2860
 Db 2182 ARNT-IEEASRMQEHIEKLETPLLHAGLFTENGDLHFLTHLVVDVAVSWRILPEFDS 2240
 QY 2861 MLY-----NGRSLPPHQSRYMYT---AD-----GRESHGFWRDVQIOTPMPTILSD 2906
 Db 2241 TAYKQAVGESIKLPKQKTSVLTYSQRIADYISISQVQREA--AYW-DECENRHQPIPK 2297
 QY 2907 DTVDGDNATKALHLSIVNIPSOVLRGSSNIITQAT-----VFNAACALVLSR 2956

Db 2298 D-----NDAASNTFKDTEVIDF--ELSRHHTELLTAHAKAYSTEMNDILLTALGLALOK 2350
 QY 2957 ESDSKDVVFGRIVSGROGL--PVEYQDIVGCTNAPVPRAHTESSDY-----NOLLHD 3007
 Db 2351 WTGNQKFLSMEGHGRSLEYLIDISRTVGTSTIYPVWLDWRSDHDKKEERLGHLLIKQ 2410
 QY 3008 IQDQYLLSLPHETIGTIGFSLKRNCTDW-----PEAITNFSCCITYHNFEYHPESQ-FEQO 3060
 Db 2411 TKDM-LHRIPKHGAGYGVLYKYSKRWGSKNSPEISFN-----LQFQDDIQSNAFEVS 2464
 QY 3061 RVEMGVLYKFNVIENDEPLYDLAIAAGEVPCAGLKVTVIAKTQLFGKRVRVHLLLEVSK 3120
 Db 2465 DIKPG--NEISPNNWERP-YALDISGAV--SSGCLNMHIYNNRFOFEERTIOTFSRHFQK 2518
 QY 3121 TFEGL 3125
 Db 2519 TLENI 2523

RESULT 8
 GRSB_BACBR
 ID GRSB_BACBR STANDARD; PRT; 4451 AA.
 AC P14588;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Gramicidin S synthetase II [Includes: ATP-dependent proline adenylyase
 (ProA) (Proline activase); ATP-dependent valine adenylyase (ValA)
 (Valine activase); ATP-dependent ornithine adenylyase (OrnA) (Ornithine
 activase); ATP-dependent leucine adenylyase (LeuA) (leucine activase)].
 GN GRSB OR GRS2.
 OS Bacillus brevis.
 OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
 OX NCBI_TaxID=1393;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 9999;
 RX MEDLINE=9221998; PubMed=1560782;
 RA Turgay K., Krause M., Marahiel M.A.;
 RT "Four homologous domains in the primary structure of GrsB are related
 to domains in a superfamily of adenylyate-forming enzymes.";
 RL Mol. Microbiol. 6:529-546(1992).
 RN [2]
 RP SEQUENCE OF 1-948 FROM N.A.
 RC STRAIN=Nagano;
 RX MEDLINE=92041751; PubMed=1939016;
 RA Hori K., Yamamoto Y., Tokita K., Saito F., Kurotsu T., Kanda M.,
 RA Okamura K., Furuyama J., Saito Y.;
 RT "The nucleotide sequence for a proline-activating domain of
 gramicidin S synthetase 2 gene from Bacillus brevis.";
 RL J. Biochem. 110:111-119(1991).
 RN [3]
 RP SEQUENCE OF 1-143 FROM N.A.
 RC STRAIN=ATCC 9999;
 RX MEDLINE=90008776; PubMed=2477357;
 RA Kraetzschmar J., Krause M., Marahiel M.A.;
 RT "Gramicidin S biosynthesis operon containing the structural genes
 grsA and grsB has an open reading frame encoding a protein homologous
 to fatty acid thioesterases.";
 RL J. Bacteriol. 171:5422-5429(1989).
 RN [4]
 RP SEQUENCE OF 1-15, AND CHARACTERIZATION.
 RC STRAIN=Nagano;
 RX MEDLINE=92011463; PubMed=1917901;
 RA Kurotsu T., Hori K., Kanda M., Saito Y.;
 RT "Characterization and location of the L-proline activating fragment
 from the multifunctional gramicidin S synthetase 2.";
 RL J. Biochem. 109:763-769(1991).
 CC !- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME, ABLE TO
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS PRO, VAL, ORN AND LEU.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC !- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOTETRAETHINES.
 CC !- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE

Db 1132 NVELOIAYSE---STEDOVERIAIEFMQFALEVRPLLRVLGVKLEAEHRLFIIMDMHHII 1188
QY 1240 SDGWSIDVLRDLNOLYSAALKSKDPLSALPLTOYSDFAKW-----OKDOFTEOEKQ 1294
Db 1189 SDGWSQMIQBIADLYE-----KELPTLGQYKDFVWHNRLLOSVDIEKQES 1239
QY 1295 L--NYWKQKOLDSPAKIPTDFARPALLSGDAGCVHVTIDGELYQSRAFCEHNTTSFV 1352
Db 1240 LAERICRDSSTESTDRLPKYOPFKALMVKDLHSQESSLWMIYTRWQKQKQOHYIWFVL 1299
QY 1353 VLLAAFRARHVLTAVEDAVIGTPIANRNRPELEDIGCFYNTQCMRINIDHDFTGTLI 1412
Db 1300 LRIMEFELS---KYGODDIWGTPIAGRSHADVENMLGMFVNTLAIRSLNNEDETFKDFL 1356
QY 1413 NOVKAATTAAFENEDIPERVVSALQPG--SRDLSSTPLAQLLFAVHSOKDLGRFRFOGLE 1471
Db 1357 ANVKOTALHAYENPDYPTDLVEKL--GIQRDLSRNPFLDFMFWLQON-TDRKSFEVEQIT 1413
QY 1472 SVP-VPSKAYTRFDEHFLQBTDSKGSVNFADLFKMETVENVVRVFFELIRNGLOSS 1530
Db 1414 ITPYVPNSRHSKFDLTLEVSEBQNETLLCLEYCTKLTDTKTVERMAGHFLQILHAIVGN- 1472
QY 1531 RPPVSILPLTDCIVTLEKLDVL---NVKHVDYPPRESSLADVFQTOVSAYPDSLAVVDSGC 1587
Db 1473 --PTIIEIIEELSEEEKOHILFEFNDTKTTPYLPCKQKDYLRNRRRRADHVAVWGKQD 1530
QY 1588 RLTYTELDRQSDILAGWLRRRRSPMAETLVAVPAPRSCETIVAFEGVLKANLAYLPDVR 1647
Db 1531 TLTYRELNERANQAVRLRQKGVQPDNIVGLLVERSPEMLVIMGILKAGGAYLPDPEY 1590
QY 1648 PSARVODILSGSGPTIYLIGHDTAPPDIEVTNVEF---VRIRDALNSNADGFEVIEHD 1704
Db 1591 PADRISYMQD--CGVRIML---TQOHLLSLVHDEFDCVILDEDSLXKGDSSNLAPVN-- 1643
QY 1705 STKPSATSLAYVLYTSGTRGPKGVMIHRVIRIIVTSGCIPNYPSETRMAHMAIADG 1764
Db 1644 ---OAGDLAYIMYTSSTGPKGVMEHRNIRLVKNTNYVQVREDDRIIQTGAIGFDA 1699
QY 1765 ASYEIYSALLFQRTLVCDVDMTTLARALKKOVFFREHVNAAHSVTSOSSODVPLRVRRLS 1824
Db 1700 LTFEFGSLHGAELPYTKVLLDAEKLHFL-----QANQIIMLTSLFL--NOLS 1751
QY 1825 RFLMEFF-----LVVTSTADPAL--DAQLYQGVQCYNGYGTENGVMSTIYPIDSTES 1877
Db 1752 QCTEEMFAGRLSLIVGGDALSFKHNNVKKCPNLTMMNGYGTENTTFTCFLD--KE 1809
QY 1878 FINGVPIGRALNNSGAYVVDPPQQLVGIGVGMELVVTGDLARGYSK--ALDENRFVHT 1936
Db 1810 YDDNIPKAIKISNSVYIMDRYQQLQPVGVGELCVGGDVGARGYMNQOPALTEERFVNP 1869
QY 1937 VNDQTVKAYRTGDRVYRIGDGLIEFFGRMDTQFKIRGNRIEABIEAALLRDSVVRDAA 1996
Db 1870 PAPGE--RMVRTGDLARW--LPDGTIEYLGRIQDQVQKIRVRIEPEGEITLLVKKHKKVESV 1927
QY 1997 VYLOQNEQAPEILGFVVADHD-----HSENDK-- 2024
Db 1928 IMVVEDNNGQKALCAIYVPEEETVSELREYIAKELPVMYPAYFVQIEQMLTQNGKVN 1987
QY 2025 -----GOSANQVE-----CQODHF----- 2038
Db 1988 RSALPKPDGEFTATYEVAPSSDIEMKLAETIWHNVLVGNKIVGLDNFFELGHSURAMTM 2047
QY 2039 -----ESGMYSIDIGEIDPST----- 2053
Db 2048 ISOVHKEFDVELPLKVLFTPTISALAQYIADGQKGYLAIOPTVPTDYYPVSSAQKRY 2107
QY 2054 IGSDPKGWTSMYDGSQIDFDE-----MHEWL-----GETTRTLHD 2088
Db 2108 ILYEPEGAGITYNPNVMFIEGKLDYQRFEYAKLSLVNRHEALRTSFYSLNCEPQVRVHQ 2167
QY 2089 NRSLL-----GNVLEIG-----TGSCMILFNLSR----- 2112
Db 2168 NVELOIAYSEAKEDEIEQIVESFVQFDFLEIAPLLRVLGVKLASDRYLFLEMDHHIISDG 2227
QY 2113 -----LESYVGLEPS-----RGAAPFVNKATESIPSL 2139
Db 2228 VSMQIITKEIADLYKGKELAEHLIOYKDFAVWQNEWFQSDALEKOKTYWLTFAEDIPVL 2287
QY 2140 -----AGK 2142
Db 2288 NLSTDYDPTIQTSEGDIVTFESAGKQABELKRLAAETGTTLYMLLLAAYNVLHKKYSQ 2347
QY 2143 AKVQVGT-----TDIGOVDDLHPD-----LVVNSVTOYPPSSEY 2178
Db 2348 EIVVGTPTIAGRSHADVENIVGMFVNTLAKKTPIAVRTFHEFLLEVKONALEAFENODY 2407
QY 2179 LAE-----IADTLHLPNV--ORIEFG-----D 2199
Db 2408 PFENLIEKLOVRDLRSRNPFLDTMFSLSDIQDOVEIGIEGLNFSPEYMOYWTAKFDISFD 2467
QY 2200 VRSQATNEHF-----LAARAHTLGKNATKDDVRQKMAEL-----EDMBE 2239
Db 2468 ILEKODDIQOFYENYCTNLFKKTETIERLATHEMHILQEIIVINPEI--KLCIEINMLSEEOQ 2525
QY 2240 ELLVE-----PAFFTSLKORFPGLVEHVELPKNMEAVNELSARY----- 2280
Db 2526 RVLYDFNGTDATYATNKIFHELFEQVEKTPDHIADVIDEREKLSYQELNAKANQALRVLR 2585
QY 2281 -----AAVHV 2286
Db 2586 QKGVPNSMVGIMVDRSLDMIVGMLGVLKAGGAYVIDIDYQERISYMWEDSGALLLT 2645
QY 2287 RGLSGDEL-----VLPVEKDDWIDFOANOLN----- 2312
Db 2646 OQKLTQOIAFSGDILYLDQEEMLHEEASNLPIARPHYIAYIITYTSGTTGPKGVMIHQ 2705
QY 2313 -----QKSLGDL-----LKSSD 2324
Db 2706 SYVNAMAMKDAYRLDTFVRLLOMASFAFAFDVSAGFARALLTGGQILVCPNEVKMDP 2765
QY 2325 AAIMAVSKIPFEITAFER--QVVASLNSNIDWQL--STIRSAEDSSLSVDPDIFRIAG 2380
Db 2766 ASLYALIK--KYDITIFEATPALVPLMEYIYQKLDISOLOLILVSGSDSCMEDFKTVS 2824
QY 2381 EAGFRVSSAFQWSONGALDAVHHCCSQGRTLNVFP--TDHHLRGSDDLTHRRPQRLQN 2439
Db 2825 REGTIRIVNSYGVTEAALILAI-----MNQPLSSLHVGT--VPICKPVANMK- 2871
QY 2440 RRIATEVRERLSSLPSYMPNSIVVLDKMLN---ANGKVDREKLSRAKVPK--- 2491
Db 2872 -----MYINQYLIQIPVGVIGELCIGGAGVARGYLNRPDLTAE--KFVNPFPV 2919
QY 2492 -----QOTAPLPTPISEVEVI--LCBEATEVEFCMKVDITDHFNFNLGGH--SLLATKLISRI 2545
Db 2920 GEKLYRTGLARMPDGNVFEFLGRNDHQVKIRGIRIELGEIEAQLRKHDSIKEATVIARE 2979
QY 2546 DQRLKVRITVKDVFHPV-----PADLASVIRQGLG-----LOQ 2579
Db 2980 DHMKYKLCAYMTEGEVNVABELRAYLANDRAAMIPSYEVSLEAMPLTANGHIDKRSLE 3039
QY 2580 P---VSDGOGQDORSAHMAPRTETEAFLCDEFAKVLGFQ--VGITDNNFLLGGHSLMATKLA 2635
Db 3040 PGCSISIGTEYDR-----PRTMLEKLEBIWKDVLGQVRGIDHDDFTTIGGHSKAMAVI 3094
QY 2636 VRIGRLDPTSVKQDVFHPVLQALALDNLVQSKTNEIVGGRMAEYSPFQLLFTEDP 2695
Db 3095 SQVHKECQTEVPLRVLFETPTIQLAKYIE-----TDT 3128
QY 2696 EEFMASETKPQLELEQII--QDIYPTOMOKAFELDHTTARPPFPVFFYIDF--PS-----T 2748
Db 3129 EOYMA-----IQPVSGQDYYPVSSAQKRFI-----VNQFVGVISYNNPMSIMLIE 3174
QY 2749 SEPDAAGLIKACESLVNHLDIERTVFAEASGELYQVVLSCDLPLOQVIEDININ---T 2804
Db 3175 GKLERTRUESAFKRIERHESLRTSFEIINGK-----PVOKTHEEVDFNMSYOV 3223

Db 2054 -----ASGVEAPRNTEETEEKLAAMVQVFLDRDRIKIGINDN 2086
Qy 2621 FFDLGGHSLMATKLAVRIGHRLDTTVSKVDVDFHVLFOALALDNLVQSTNETVGGRE 2680
Db 2087 FFEIGGDSIKALQIVSKLS-RADLKQLQVLDLFTNPFIRHLSKYVKKETKARTSEIVQQ- 2144
Qy 2691 MAEYSPFOLLTFEPEFMASEIKPQLELOEIIQDIYPTOMQKAFLEFDHTTARPPFPV 2740
Db 2145 -VPLTPVQORSE-----PEANQREQNH-----NOAFML-----YRENGFAE 2179
Qy 2741 FYIDPSTSEDAAGLAKACSLVNLHLDIFRTVFAEASGEIYO-----VLSCLDLPQVI 2796
Db 2180 RIVE-----KVFRKLTEHHDALFVWYWEKNGDIIQNRGLDESVDLYVYDL 2226
Qy 2797 ETEENINTATNEFLDEFAPVRLCHPLIRITIKOTKSMRVIMRISHALYDGLSLEHV 2856
Db 2227 KTERNLEKTVQIATNTQKDISISSEKMKLCVPTTEGDHLLTAIHLHLLVDGVSWRILF 2286
Qy 2857 RKLHLY-----NGRSLPPHPSRYMQYTAGRESGHH-----FWRDVIQNTPTMIL 2904
Db 2287 EDFEAAAYQALQGRPIELGYKTSYKTFSEKLAEVANSKLLKEQEVYREISKGMFLP 2346
Qy 2905 SDDTWDGNDATCALKHLSKIVNPISOVLRCSSNII-TQATVFNACALVLSRESKDV 2963
Db 2347 KHROAHDNYENSRTLSISQTEQTEQLLKEAHKAYNTQINDLLLTALLIASRQLTGENR 2406
Qy 2964 VFGRIVSGRQLPVEYQDI-----VGPCTNAPVRAHIESDYDQQLLDHIDQYLLSLPH 3018
Db 2407 L--KILMBEGHRRDILQDVDTITRVGWFMTAMPYFVIDLEDEADLSVMKIKVETLRKIPN 2464
Qy 3019 ETIGFSDLK 3027
Db 2465 NGIGYGILK 2473

RESULT 10
HTSL_COCCA STANDARD; PRT; 5217 AA.
AC Q01886;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE HC-toxin synthetase (EC 6.3.2.-) (HTS).
GN HTSL.
OS Cochliobolus carbonum (Bipolaris zeicola).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=5017;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 90305 / SB111;
RC MEDLINE=93100328; PubMed=1281482;
RA Scott-Craig J.S., Panaccione D.G., Pocard J.-A., Walton J.D.;
RT "The cyclic peptide synthetase catalyzing HC-toxin production in the
RT filamentous fungus Cochliobolus carbonum is encoded by a
RT 15.7-kilobase open reading frame."
RL J. Biol. Chem. 267:26044-26049 (1992).
RN [2]
RP FUNCTION.
RC STRAIN=ATCC 90305 / SB111;
RC MEDLINE=20138231; PubMed=10671527;
RA Cheng Y.-O., Walton J.D.;
RT "A eukaryotic alanine racemase gene involved in cyclic peptide
RT biosynthesis."
RC J. Biol. Chem. 275:4906-4911 (2000).
CC -!- FUNCTION: Non-ribosomal peptide synthetase, able to activate
CC proline and AEO (2-amino-9,10-epoxi-8-oxodecanoic acid), and
CC epimerize L-Pro. Catalyzes the production of HC-toxin: a cyclic
CC tetrapeptide. Activates and thioesterifies L-Pro, and epimerizes
CC it to D-Pro; also uses D-Ala as a substrate but this is epimerized
CC from L-Ala by TOXG.
CC -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.

CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M98024; AAA33023.1; .
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00550; pp-binding; 4.
DR Pfam; PF00668; Condensation; 5.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
DR PROSITE; PS00455; AMP_BINDING; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
KW Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
FT REPEAT 249 842 DOMAIN 1.
FT REPEAT 1854 2452 DOMAIN 2.
FT REPEAT 3006 3606 DOMAIN 3.
FT REPEAT 4158 4738 DOMAIN 4.
FT DOMAIN 762 840 ACYL CARRIER (ACP) 1.
FT DOMAIN 2384 2450 ACYL CARRIER (ACP) 2.
FT DOMAIN 3536 3604 ACYL CARRIER (ACP) 3.
FT DOMAIN 4667 4736 ACYL CARRIER (ACP) 4.
FT BINDING 803 803 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2414 2414 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3568 3568 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4700 4700 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 5217 AA; 574637 MW; 0331D9C5400163A5 CRC64;

Query Match 10.18; Score 1625.5; DB 1; Length 5217;
Best Local Similarity 22.38; Pred. No. 3.2e-85;
Matches 736; Conservative 475; Mismatches 1080; Indels 1007; Gaps 125;

Qy 46 IEALKPCTPFOLDMIDCNALDKQSA-----ICHAVIDVPTDIDISRFALAKWEIYNQ 97
Db 1337 LEGYVPGSPQDALFLSQSKSQDCGAYEVDFTWRVATSLQNSQPAVDIGCLVANKDTVAL 1396
Qy 98 TPALRAFATSD---SGKTSQVILKDSFVFSMMCMWSSSSSPDEV---VRDEAAAA---A 147
Db 1397 HAALRTVILESSLPATGILHGVLR-----SHDPDIDVLDVRDVTAAITLDS 1444
Qy 148 SGPCRNFEVLLEDQTKKCOLVWTF-----SHALVDVTFQQRVLSRVFAAYKHEK 197
Db 1445 YPPTTEGIALIKRPPHRLICTTIEGSLVKFKVNHLLVFDGMDTKTIQDLSKAYTCRH 1504
Qy 198 DTHRPETPSS--DAT-----DTSQSVSVSVSMSCEDNAVSATHFQW 237
Db 1505 SNKLDPHSESKLHDGTGNGRPTKPLAEFIRYIRDPQKQDSI-----NYWK 1551
Qy 238 THLNDLNASVFPPLHSDHL-----MVPNTTAEHRTFPPLSQKALSATCRT 285
Db 1552 NALRGATTCSPPPLFDQITSEKAMPROSASWVPIPLCVDSKELSKTLANLGTMTSTMFOT 1611
Qy 286 ALSTLLSRVTHSDEALFCAVTE--OSLPDPKHLYADGYQTVAPLRVHCQSNLRASDVMDA 344
Db 1612 VWALVLIYSONGSQSVFGYLFSGRDAPVDGSDSANGF--IA-----MLVC 1655
Qy 345 ISSYDDRLGHLAPFLRLDIRNTGNGSA--AC----DFQTVLLVTDGSHVNGINGFLQOI 399
Db 1656 FFDFDDGVTIVADMARKIHNASANSISHQACSLAEIQDALGLSTSTPLFTATYLPK- 1714

QY 400 TESSHEMPCNRA-----LLHCOMESSGALLVAYDHNV--ID 436
Db 1715 -----RPNVKGPEHLCFEELSMDSPTEFDLTLFVEPTQESNEVSAHLDFKLSYIS 1768
QY 437 SLOTRLLQOFGHLI-KLQSPDLJLSSMAEVLNMTEDYDRAEIESWNSQPLEVQDTLIHHE 495
Db 1769 QAYATSIATVAHLSLVHDP--YRALNTLPVSEHDTAIRSNDHLPFPATECIEHET 1826
QY 496 MLKAVSHSPKTAIQAWGDWYSELNYSRLAVHIKSLGRARAOAIIIPVPEKSKWVI 555
Db 1827 FSKVVEHPQREACISWGDLSYAELSLSQRLSILHLSLGIKVGT-K-IPICEKSMWTI 1885
QY 556 ASMLAVLKSGNAFTLDPNDPAPTAQVVTQTRATVAL-----TSKLHR-ETVOKLVG--- 607
Db 1886 VTILAVQAGGVFVLEPGHPSRSGIIGQVQAEILLCSPTSRMGALQNTSTOMGTEF 1945
QY 608 RCVVODELLQSS--VSASDDFSLTKSQDLAVYFTSGTGDPKGMIEHRAFPSSCALAF 665
Db 1946 KIVELEPEFIRSLPLPKPHQPMVGLNDDLYVFTSGTGVPKGAATHQVATGIYEH 2005
QY 666 GASLGINS---DTRALQFTHAFGACLEIMTTLINGGCVCIPSDDDRMNS-IPSFINKY 721
Db 2006 AVACGMTSLGAPRSLOFASYSFASIGDIFTTAVGGCLCIPREEDRNPAGITTFINRY 2065
QY 722 NYNNMATPSYKCTFSPEDVPGLATLVLYGEOMSSSVNAILWAPKLOLLNGYQSESSIC 781
Db 2066 GVTWAGITPSLALHLPDPAVPTLKLCVAGEPLSMVSVVWMSKRLNLINNYGTEATVAC 2125
QY 782 FASNM---STEPNMGRVAGHNSWIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPP 838
Db 2126 IANOVCTTIVTSDIGRGYRATVWVQPDHNSLSVPIGAVGELIIEGSLCRGYL-NDPE 2184
QY 839 EKSPFTDTPSWYPANTFPDGAKLYRTGDLARYASDGSIVCLGRIDSQVKIRGORVELGA 898
Db 2185 RTAEVETIRPSW--LHDLRPNSTLYKTGLVRYSDGKIIFIGRKDTQVKMNGQRELFGE 2242
QY 899 IETHLRQOM-PDDLTVVATKRSQSANSTSLIAFL-IGSSVFGNRPSPDAHILD----- 950
Db 2243 VEHALQLOLDPDGPPIVDLLKRTQSGEPDLTIAFLFVGRANTGNSDEIFATSTSSL 2302
QY 951 -----HDATKAINIKLEQVLRHSPISFYICMLE--LPRTATGIDRRRLRIM-- 996
Db 2303 SEESTVIKKLODAQRAM-----EVLPLFWVQYAI-PIEGGIPLTAAKIDRMLKICE 2356
QY 997 --GKDLDKQTOGAI---VQAPAPIPVFAFATAA-KLHSIWQSLGIDPATVNVG--ATF 1048
Db 2357 PFNRNDLISFTSKALSTSVDKDAET-----TDTVEDRLARIWEKVLGVK---GVGRESDF 2407
QY 1049 FELGNSITAIKWNVNWARSVGMDLKVSNIYQHPTLAGISAVVKG-----DPLSYTL---IP 1101
Db 2408 FSSGNSMAAIALRAEPAQRSGFTLFVADIFTNPRLADMAKLFSGHQSVPSSSTLRKVP 2467
QY 1102 KSTHEG-----PVEQSYSGO-----RWFLDQ 1123
Db 2468 ISSLOKRSGLQATAPVNSGVPVRCOKENIDCPVAFYEBEGSPDTQLKEASRICGISS 2527
QY 1124 LDVGSWYLIPIAVMRGPNVVDALRRALAALE-----QRHETLRTTFEDQDQV----- 1172
Db 2528 RSIEDVFPCTP-----MOEALVALSLIPGAQASYALHAFAELRPGLDNRFR 2574
QY 1173 -----GVQIVHEKLEEMKVIDLQSGDLDLPFVLNQBQETTPEN 1210
Db 2575 SAWESTVKAQPIILRSRIISGSSGVVTSATDSIPQLDVG--LDTF--LEQLOQVGFA 2630
QY 1211 LSSEAGWRATLLRGLGDDHILITVMHHIISDGWSIDVLRRLDNLNOLYSAALKDSDPLSAL 1270
Db 2631 PCAPLERLAFVYSKADDCDYFVLSAHHAIDYDGSNLNLSWSQVLYATNG----- 2679
QY 1271 TPLPTOYSDFAKWQDKQFTEQBK--OLNYYKKOL--KDSPPAKIP-----TD 1313
Db 2680 -ELPPPGPSFKHARNLNLVQSKLSDSEDFWRKLLVKPDQESFRFPDVPVGHKPAIRCTTN 2738
QY 1314 FARPALLSDAGCVHVTIDGELYQSLRAFACNEHNTTSFVLLAFAAAHYRLTAVEDAVI 1373

Db 2739 FHEPFSMQSKIG-----TTANTCINAAMAITTAAQYSSNKTVMF 2776
QY 1374 GTPIANRNRD--ELEDIIGCFVNTQCMRINIDHHDFTGLINOVKATTTAAAFENEDIPFE 1431
Db 2777 GVTLWGRDFPMIDIEHTGPTIVTVPROVNVIPESSVAEFLQDLQKSLAVVLPHOHLGLH 2836
QY 1432 RVYSALQPSGR---DLSSTPLAQLIFAVHSOKDLGRFKQGLSVPPVPSKAYTRFDMFEH 1488
Db 2837 R-IQALGPTARQACDFST-----LLVNHGSSISWSELEAADIVPPLKSSDLAYPMV 2889
QY 1489 LFQE---TDSLKSVNFADELFMETVENVVRVFFELLRNGLOSSRTPVSIPL-----L 1539
Db 2890 VEVENASSDTLDIRVHSDPDCEIVQLLERLMEQF---GHNLOTLCLRAASFDEGKRITAE 2945
QY 1540 TDCIVT--LEKLDVLANVKHVDYPRESSLA--DVFOQTQSVAYPDSLAVVDSSCTTYTELD 1595
Db 2946 MDDTATHTLTLESWNSRVKSDPVAIAIAVHKLEETAQSPAESAIAVHGDQLSYMQMD 3005
QY 1596 ROSDILAGMLRRSRM--PAETLVAVFAPRSCETIVAFFGVGLKANLAYLPDVRSPSARVQ 1653
Db 3006 RCADVLARQIRKTNMISAQSPFVCIHLLRSATAVVSMVLAVKAGGAMPVDISQPSRLQ 3065
QY 1654 DILSGLSGPTIVLIGHDTAPPD-----TEVTNVEVRIRDALNDSNADGFVEIEH 1703
Db 3066 NLIEE-SGAKLVL---TLPEANALATLSGLTKVIPVLSSELVQQITDNTTKKDEYCKS 3120
QY 1704 DSTKPSATSLAVVLYTSGSTGRPKGVMIHRVIRTVTSGCIPNY---PSETMAHMATI 1760
Db 3121 GQTDPS--SPAYLLYTSGTGPKGVVMEHRANSLGFT--CHAEYMGFNCTRLILQFSSL 3176
QY 1761 AFDGASYEISALLFGRTLVCVDYMTLDARALKDVFREHVNAAHSVTSSSODVPLRVP 1820
Db 3177 MFDLSILEIWAIVLYAGCL-----FIPSDKERYNNLQDFTINDINTVLTPTSGKLLNP 3231
QY 1821 RRLSRLTLMFFLVVVTSTAPDALDAQLYGVOCYNGYGTENGVMWSTIYPIIDSTESFIN 1880
Db 3232 KDLPN--ISFAGPIGEPMTSLIDAWTL-PGRRLVNSYGPTEACVLITAREISPTAP--H 3286
QY 1881 GVP---IGRALNNSGAYVVDPEQ-QLVGVIGVMGELVVVTGDLARGY--SDKALDENRVRHI 1935
Db 3287 DKPSSNIGHAL-CANITWVVEPQRTALVPVIGAVGELCIEAPSLARCILANPERTEYSEPT 3345
QY 1936 TV-NDQTVK---AYRTGDRVYRIGDGLIEFFGFRMDTQFKIRGNRIESAIEAALLRDS 1991
Db 3346 VLDNMWTKKGRVYRTGDLVRY-ASDGTLDLFGKDKGQIKLRGORTLGEIE----- 3396
QY 1992 VRDAAVVLQONEDQAEIILGFVVADHDHSENDKGSANQVGEHQDHIFESGNSYSDIGEIDP 2051
Db 3397 -----HHIRLMSD-----DP 3407
QY 2052 STIGSDFKGTSMYDGSQIDFDBMHEWLGETTTLHNRSLGNVLEIGTCGMLFLNLS 2111
Db 3408 -----RFHE----- 3411
QY 2112 RLSEYVGLPEFSAAAFVNKATESIPSLAGAKVQV-GTATDQGVDDLHPDLVLVNSVI 2170
Db 3412 -----ASVQLYNPATD-----PD----- 3424
QY 2171 QYFPPSEYLAETADTLIHLPNVORIFFGVRSQATNEHFLAARAIHTGKNATKDDVRQK 2230
Db 3425 -----RDATVD----- 3430
QY 2231 MAELEDMEBELLVEPAFFTSLKDRFPGLVEHVEILPKNMEAVNLSAYRYAAVVHVRGSL 2290
Db 3431 ----- 3430
QY 2291 GDELVLPEKDWIDFQANQLNOKSLGDLKSSDAAIMAVSKIPFEITAFERQOVASLNS 2350
Db 3431 -----VQMPREPLYLAGL----- 3442
QY 2351 NIDEWOLSTIRSAEGDSSLSVPDIFRIAGEAGFRVEVSAROWSANGALDAVFHCCSQ 2410

```

Db 3443 -----VLDL-----VSLMRSDS-----MHVNI 3461
QY 2411 GFLVNFDPDHLRGSDLLTNRLQRLQNRRIAIEVRRLRSLLPSYMPISNIVVLKMP 2470
Db 3462 ANISENLQT-----LVTELKSLRGVLPHYMWPLHFVAVSRUP 3499
QY 2471 LNANGKVDKELSRRAKVPKQTAAPLP-TFFPISSEVILCEEATEVFGMKVDIITDHF 2529
Db 3500 TGSSGKLDHAFV--RACL--RELTAFLDGNFP--KVEQVLTN-----3536
QY 2530 NLGGHSLATKLISRIDQRKVRITVKDVEDHVFADLASVIRGLG--LOOPVSDGG 2586
Db 3537 -----ESVLKQWGTVLAMDPHSIGIQG 3558
QY 2587 QDRSAHMAPRTEFAILCDEFAKVLGFQVGIITDNFELGGHSLMARTK---LAVRIGHRLD 2643
Db 3559 -----DDFESLGSSISAMRLVGLARSSGHKLQ 3586
QY 2644 TTVSVKDVDPHVPVLFOLATALONLVOSKINEIVGGREMAEYSP-----FQLLTEDPE 2696
Db 3587 H-----EDIFMCPRLADMAGQI-SFVQE-----ASVSPTTSPTIKFDLLDCEVD 3630
QY 2697 EFMAEIKPOLEL-QEIIQDIYVSTOMQAKAFLDHTTARPRPVPFYIDFPSTSEPDAG 2755
Db 3631 E-VIDHLPOLDMKNKELIEDVPTPLQESLM--AATARHGEAYTM-IOSTIVLAAQLAQ 3686
QY 2756 LKACSLNVHLDIFRTVFAEASGELYQVVLSCLDLPIQVIEFDNIN----TATNEFLD 2811
Db 3687 LKAMDWFRDFEVLRT-----RIALGPSQALQVVKHEELSWESFPSIQSFKD 3736
QY 2812 EFAPKPVRLGHPLIRITIKQ-----TK-----SMVIMRISHALVDGLS 2851
Db 3737 HFYRS-LGVGKPLARLAVITQALDKOPTISHTREARTKNSQDTVMVVGCAHHSIYDAH 3795
QY 2852 LEHVVRKLHMYNG-----RSLPPHOFYSRYMOVTADGRESHG--FWROV 2895
Db 3796 LSMWRRLYREFIGSQADGILEAETSRSSEGVV---FKSYVEKLLRGKDNDESLLEWKEK 3852
QY 2896 ----IQNPTWILSDDTVVDGNDATCKALHLSKIVNIPSOVLRGSSNIITQATVFNACAL 2952
Db 3853 LRGVSSSQPPASWPRVLEHQPFSATQTL-ITR-VSLPTSSRKLG--ATVATVAYAAWAL 3908
QY 2953 VLSRESDSKDVFGRTVSGRO---GLPVEYQDIVGCTNAPVRAHIESSDYNQLLHD 3007
Db 3909 TIAHYTADPDVFGATLSGRETMAGSISHPESIAGTPIITVPLRIII----DQTVVSD 3963

RESULT 11
SRFL_BACSU
ID SRFL_BACSU STANDARD; PRT; 3587 AA.
AC P2706;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Surfactin synthetase subunit 1.
GN SF6A OR SRFA OR SRFA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC STRAIN=168;
RX MEDLINE=93181186; PubMed=8441623;
RA Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RT Zuber P., Yamane K.;
RT "Nucleotide sequence of 5' portion of srfA that contains the region
RL required for competence establishment in Bacillus subtilis.";
RN Nucleic Acids Res. 21:93-97(1993).
RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN=168 / JH642;
RC MEDLINE=93360813; PubMed=83355609;
RA Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,

```

```

RA Venema G., van Sinderen D.;
RT "Sequence and analysis of the genetic locus responsible for surfactin
RT synthesis in Bacillus subtilis.";
RL Mol. Microbiol. 8:821-831(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
RT chromosome: determination of the sequence of a 146 kb segment and
RT identification of 113 genes.";
RL Microbiology 142:3047-3056(1996).
RN [4]
RP SEQUENCE OF 1-460 FROM N.A.
RX MEDLINE=91154134; PubMed=1847909;
RA Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
RA Zuber P.;
RT "srfA is an operon required for surfactin production, competence
RT development, and efficient sporulation in Bacillus subtilis.";
RL J. Bacteriol. 173:1770-1778(1991).
RN [5]
RP SEQUENCE OF 1-38 FROM N.A.
RX MEDLINE=91358326; PubMed=1715856;
RA Nakano M.M., Xia L., Zuber P.;
RT "Transcription initiation region of the srfA operon, which is
RT controlled by the comp-comA signal transduction system in Bacillus
RT subtilis.";
RL J. Bacteriol. 173:5487-5493(1991).
RN [6]
RP SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=168;
RX MEDLINE=95219080; PubMed=7704255;
RA Fujishima Y., Yamane K.;
RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
RL of srfA of the Bacillus subtilis chromosome.";
RN Microbiology 141:277-279(1995).
CC -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13262; BAA02522.1; -
CC EMBL; X70356; CAA49816.1; -
CC EMBL; D50453; BAA08982.1; -
CC EMBL; M59939; AAA22815.1; -
CC EMBL; M64702; AAA22816.1; -
CC EMBL; D30762; BAA21034.1; -
CC EMBL; Z99105; CAB12142.1; -
CC FIR; S35517; S35517.
CC HSSP; P14687; IAMU.
CC Subtilist; BG10168; srfAA.
CC InterPro; IPR000873; AMP-bind.
CC InterPro; IPR001242; Condensatn.
CC InterPro; IPR003880; Ppantne_attach.
CC Pfam; PF00501; AMP-binding; 3.
CC Pfam; PF00550; pp-binding; 3.
CC Pfam; PF00668; Condensation; 4.
CC PRINTS; PR00154; AMPBINDING.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
CC PROSITE; PS00455; AMP_BINDING; 3.

```

DR PROSITE: PS50075; ACP_DOMAIN; 3.
KW Ligase; Antibiotic biosynthesis; Complete proteome.
KW Multifunctional enzyme; Repeat; Domain 1 (GLU-ACTIVATING).
FT REPEAT ? 1047 DOMAIN 1 (GLU-ACTIVATING).
FT REPEAT ? 2086 DOMAIN 2 (LEU-ACTIVATING).
FT REPEAT ? 3114 DOMAIN 3 (D-LEU-ACTIVATING).
FT DOMAIN 976 ACYL CARRIER (ACP) 1.
FT DOMAIN 2015 ACYL CARRIER (ACP) 2.
FT DOMAIN 3043 ACYL CARRIER (ACP) 3.
FT BINDING 1006 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 2045 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 3073 PHOSPHOPANTHETHEINE (POTENTIAL).
FT CONFLICT 146 I -> M (IN REF. 2 AND 3).
FT CONFLICT 151 L -> V (IN REF. 2 AND 3).
FT CONFLICT 165 A -> G (IN REF. 2, 3 AND 4).
FT CONFLICT 281 T -> Q (IN REF. 2 AND 3).
FT CONFLICT 460 D -> T (IN REF. 3 AND 3).
FT CONFLICT 540 P -> A (IN REF. 2 AND 3).
FT CONFLICT 562 I -> Y (IN REF. 2 AND 3).
FT CONFLICT 639 GS -> PT (IN REF. 2 AND 3).
FT CONFLICT 644 R -> P (IN REF. 2 AND 3).
FT CONFLICT 647 L -> F (IN REF. 2 AND 3).
FT CONFLICT 649 A -> P (IN REF. 2 AND 3).
FT CONFLICT 1026 L -> Q (IN REF. 2 AND 3).
FT CONFLICT 1065 SLAQQTYIVSDFDAGVGNPAAALEGPDIDQKLERAF
QGLIRRHSLR -> HWSSEHSSASSMRRESATHASSN
SGGFRYSKAGARISGNTPTPRVIE (IN REF. 2 AND 3).
DS -> VC (IN REF. 2 AND 3).
R -> A (IN REF. 2 AND 3).
V -> L (IN REF. 2 AND 3).
S -> T (IN REF. 2 AND 3).
D -> T (IN REF. 2 AND 3).
R -> A (IN REF. 2 AND 3).
LRLCSKWT -> PAVFIQMD (IN REF. 2 AND 3).
L -> P (IN REF. 2 AND 3).
H -> Q (IN REF. 2 AND 3).
QQ -> HR (IN REF. 2 AND 3).
V -> L (IN REF. 2 AND 3).
S -> C (IN REF. 2 AND 3).
R -> S (IN REF. 2 AND 3).
RC -> AV (IN REF. 2 AND 3).
E -> ENPE (IN REF. 2 AND 3).
T -> S (IN REF. 2 AND 3).
SP -> TA (IN REF. 2 AND 3).
P -> N (IN REF. 2 AND 3).
F -> N (IN REF. 2 AND 3).
R -> S (IN REF. 2 AND 3).
Y -> S (IN REF. 2 AND 3).
DE -> HQ (IN REF. 2 AND 3).
DAGL -> HPPF (IN REF. 2 AND 3).
R -> T (IN REF. 2 AND 3).
E -> H (IN REF. 2 AND 3).
GQ -> PH (IN REF. 2 AND 3).
R -> E (IN REF. 2 AND 3).
R -> E (IN REF. 2 AND 3).
SEQUENCE 3587 AA; ALE2DABFF93EDE3A CRC64;
Query Match 10.0%; Score 1613; DB 1; Length 3587;
Best Local Similarity 21.6%; Pred. No. 8.9e-85;
Matches 696; Conservative 483; Mismatches 1101; Indels 946; Gaps 108;
QY 82 IDSRPALAWKEIVNOTPALRAFAFTSDSGKTSQVI-----LKDS-----FVF 124
DB 42 IDVLVEQAIQEFIRNDAMRLRLDENGEPVQVISEYRVPVDIKHTDITDPNAIEFTS 101
QY 125 SHMCSSSSSPDEVVDEAAAAAGPRCN--RFVLLDMOTKKCOLVWTFSS---HALVDV 179
DB 102 QMS-----REETKPKPLPLYDCDLFRFSLF---TIKENVWFYANVHVHVIDSG 145
QY 180 TFQORVLSRVFAAYKHEKDTHRPETPSSDARDTDSQSVSVVSMSCEDNAVSAATH----- 234
DB 146 ISWNILGNAMIMHYL-----ELASASETKEGISHSHFIDHVLSEQYQAQSRPEKOKA 197

QY 235 FWOTHNLNDL-----NASVFPHL--SDHLMVNPNTTAEHRITTEFLPSQKALSNSAICRT 285
DB 198 FNNKQFESVPELVSLKRNASAGSGLDAERFSKDVPEALHQIILSPCEANK--VSVLSVFQS 256
QY 286 ALSILLSRTHSDEALFGA-VTEQSLPFDKHYLADGTOTVAPLPRVHCOSNLRASD-VMD 343
DB 257 LLAAYLYRVSGONDVVVTGTFMGNRTNAKEQKML--GMFVSTVPLRNTIDGGOAFSEFVKD 314
QY 344 ATSSYDDRLGHLA--PGL--RDIRNTGDN-----GSAACDFOT-----VLLVTDGSHV 368
DB 315 RKMDLMKTLRHOKYPYNLLINDRETCKSLTKLFTVLSLEYVQMOWKEEDLAPLETPIFS 374
QY 389 NNINGFLOQITESHFPCNNRALLHQCMESSGALLVAYYDHNVIDSLQTLRLQOFG 448
DB 375 GSGLND-----VSIH-----VKRDWDTGKLTIDED 399
QY 449 HLKICLOSPDLSSMAEVNLMTEYDRAETESNNSQPLEVOD--TLI---HHEML----- 497
DB 400 Y-----RTDLFSREEINNICERMITMLENALTHTPEHTIDELTISDAEKEKLLARAG 452
QY 498 KAVSHSPTKT-----AIQAWDGDWTYSELDNVSSSLAVHHSIGSLGURAQ 541
DB 453 KVSYHKDMTIFELFOEKABLLSDHPAVVFEDRTLSTRLHQSARIANVLOKQGV--GPD 511
QY 542 AIIPVYFEKSWVIASMLAVLKSNGAFTLIDPNDPPARTAQVVTQTRATVALT--SKLHR-- 599
DB 512 SPVAVLIERSERMITAIMILKAGGAYVPIDPGFAERIQYILEDCGADFILTESVAAP 571
QY 600 ETQVKLVGRVVDDELLOSVSASDDFSSLTQSODLAYVIFTSQDPSKPGIMIEHRAFS 559
DB 572 EADAELIDIDQAEIEGAEESLNAD-----VNARNLAYIYTSQGTGPRKGVMIHQVH 625
QY 660 SCALRFGASLIGNSDTRALQFGTHA---FGACLEITMTTLINGCVCIFSDDDDMN--SI 714
DB 626 HLVESLQOYI--YQSGSQTLMALLAPFHDASVKQIFASLLIGQTLIYVPKKTVTNGAAL 684
QY 715 PSFIRNYNNMMATPSYMGTFSP--DVPLAT--LVLYGEOQSSSVNAIWAPKL----- 766
DB 685 TAYRKNSTEATDGTFAHQLMAAAGDFEGLKHKMLIGGEGLSVV---ADKLLKLFK 740
QY 767 -----QLLINGYQSESSSICF--ASNKSTEPNN-----MCRAYGAHS--WVIDPNDIN 810
DB 741 EAGTAPRLTNVYGPTE---CVDASVHPVIPENAVQSAVYPIGKALGNRLYLDDQK--G 795
QY 811 RLVPIGAVGELVIESPGIARDYIVPPPPEKSPFETDIPSWYPANTEPDGAKLYRTGDAR 870
DB 796 RLQPEGVAGELYIAGDVGVRGYLHLPELTEEFLQD-----PVPDGRMYRTGDVVR 847
QY 871 YASDGSIVCLGRIDSQVKIRGORVELGAIETHLRQOQMPDGLTIVVEATKRKOSANSTSLI 930
DB 848 WLPDGTIEYLGREDQVKVRYRIELGEIAVI--QAPDPAKAVVLA--RPDQGNLEVC 904
QY 931 AFLIGSSYFCNRPNSDAHILDHDTAKAINKLEQVLRPHSPISFYICMLLELPRATGKIDR 990
DB 905 AYVWQKP--GSEFAPAGLREHAARQ-----LPDYMVPAYFTEVTEIPLTPSKVDR 953
QY 991 RRLRMKGDKILQTOGAIVQOAPAPIPVFADTAALKHSIWOSLGIDPATVNVGATFFE 1050
DB 954 RKL-----FALEKAVSGTAYTAPR-----NETEKAIAALWQDVLNVEKA--GIPDNFFE 1001
QY 1051 LGGNSITATKMV--NMARSVGMDBLKVSNIYOHPTLAGISAVV--KGDPLSYTLIPKSPHEGP 1108
DB 1002 TGGHSLKAMTLLTKIHKTGIEIPLQFLFEHTITALEAEADHRESKAFAVIEPAEKQEH 1061
QY 1109 VEGSYVSGRLWFLDQDLVCSLWYLPYAVMRGPNVDALRALAALORHETLRTTFED 1168
DB 1062 YPLSLAQOQRTYIVSOFEDAGVGNMPAAAILEGPDIDQKLERAFQGLIRHESLRTSFLV 1121
QY 1169 QDGVGVQIVHEKLSKEMKVIDLCG--SDLPDFEVLNQEQTPFNLSSEAGWRATLRLRGED 1227
DB 1122 ENSTPRQKTHDSVDNFIEMIERGGRSD-----EAIMASFVRTFDLAKRPLFRIGLLGLEEN 1177

QY 1228 DHILITVMHHIISDGSIDVLRDLNOLYSAAALKDSKDPLSALTPLPIQYSDFAKWQKQ 1287
Db 1178 RHMLFDMHHLISDGVSIGIMLEELARIYK-----EQPLDLRLQYKDVAVMQSRQ 1228
QY 1288 FIE-OEKOLNWKOLKSDSPA-KIPTDFARPALLSGDACCVHVTIDIGLYQSLRAFCE 1345
Db 1229 AAGYKQKQAYWEVAFAGELPVQLLSYDPRPPVQSFEGDRVSIKLDAGVKDRNLRLAEQ 1288
QY 1346 HNTTSFVLLAFAAHRYLTAVEDAVIGTPIANRNRPELEDIIGCFVNTQCRINIDHH 1405
Db 1289 NGATLYMVLSAYITLLSKYTGQDDIIVGTPSAGRNHSDTEGIGMFVNTLAIRSEVKQN 1348
QY 1406 DTGCTLINQKATTAFAFENEDIPFERVVSALQPSRDLSTPLAQILFAVHSQKD----- 1461
Db 1349 ETTQLISVRKRVLDAFSDHODYPFEWLVEDLNI-PRDVSRRHPLFTDMFSLQNATEGIPA 1407
QY 1462 LGREFKQGLSVVPKAYFRFONEFHLFOETSLKGSVNFADFLFKMETVENVVRVFE 1521
Db 1408 VGDLSL-----SVOETNEFKIAKFDLTVOARETDEGIEDVDYSYKLFKQSTADRLVYHFA 1463
QY 1522 ILRNGLOSSRTPVSIPLPTDGIIVTLEKLDVLYNKHVDYIPRESSLADVFOTQVAYSPLSA 1581
Db 1464 LLEDAADPEKPISEYKLLSEEAASQIQOFNPGRTYPKDKTIVQLFEQAANTPDHTA 1523
QY 1582 VDSSELTYTELDRQSDIILAGWLRRRSMPAETLVAVFAPRSCETIVAFEGVLKANLAYL 1641
Db 1524 LOYEGESLTYRELNRANRLARILSLGAGEGRTAAVLCERSMDMIVSILAVLKSGSAYV 1583
QY 1642 PLDVRGPSARVQDILSGSGTIVLIGHDTAPPDIEVTNVEFVRIRDALNDSNADGFEVI 1701
Db 1584 PIDPEPIQRMQHFRD-SGAKVLLTOR-----KLKALAEAEFKGVIVL 1627
QY 1702 --EHDSTKPSATSLAYVL-----YTSGSTGRPKGMVIEHRVIRTWTSCIPNYP 1750
Db 1628 ADEESYHADARNLALPLDSAAANLTYTSGTGTGPKGNVTHANILRTVKE---TNYLS 1684
QY 1751 ETR--MAHMATAFQASVEIYSALLFRTLVCVDYMTLDRALKDVFREHVNAASH 1807
Db 1685 ITSQDTILGSNVVDFAMFDMGSELLNGAKVLLIPKETVLDMARLSRVTERENISILMI 1744
QY 1808 VTSSQDVLVRPRRLS--RTLMEFFLVVTDSTAPDALDAQGLYQGVQYNGYGTENG 1865
Db 1745 TTAFLHLLVDLNPACLTSLRKIMFGGERASVEHVRKALQTVG--KG-KLLHMYGPSESTV 1801
QY 1866 MSTIYPIDSTESINGVPIGRALNNGAYVVDPEQOLVGTVGMGELVWTGDLARGYSDK 1925
Db 1802 FATYHPVDELEHTLSVPICKPVSNTEVYILDRTHGVQVAGIAGELCVSGEGLVKGYNR 1861
QY 1926 -ALDENRFV-HITVNDQTKAYTRDVRVYRIGDGLIEFFGRMDTOFKIRGNRIESAEIE 1983
Db 1862 PELTEKFPVPHPTSGE--RWYKTGLARW-LPNGDIEFGRIDHVKIRGQRIELGEIE 1918
QY 1984 AALLRDSVRAAAVVLQONDDQAPEILGFYVADHDHSENDKGSANQVQGWQDHFESGM 2043
Db 1919 HQL-----QTHDRVOESV----- 1931
QY 2044 SDIGEIDPSTIGSDFKGTWSDYGSQIDFDEMHEWLGETTRTLHDNRSLGNVLEIGTSG 2103
Db 1932 -----VLAVDQAG 1940
QY 2104 MILFNLDRLSEYVGLSPRSAAAFVNKATESIPSLAGKAKVQVGTATDIGVDLDLHPDL 2163
Db 1941 -----DKLLCAY----- 1947
QY 2164 VLVNSVIQPPSSEYLAETIADTLIHLPNVORIFFGDRVSQATNEHFLAARAIHTLGNAT 2223
Db 1948 -----YVGE-----GDISSQEMREH-----AA 1964
QY 2224 KDVROKMAELEDMEELLVEPAFTSLKDRFPGLVEHVEILPKNEAVNELSAYRYAAV 2283
Db 1965 KD----- 1966
QY 2284 VHRVSGLDGLVLPVEKDOWIDFQANQLNOKSLGDLKSSDAAIMAVSKIPFEITAFERQ 2343

Db 1967 ----- 1966
QY 2344 VWASLNSIDEWQLSTIRSSAEGDSSLSVPDIFRIAGEAGFRVEVSARQWSONGALDAV 2403
Db 1967 ----- 1966
QY 2404 FHHCCSOGRTLNVNPTDHHLRGSDLLTNRPLQRLQNRRIAEVRERLRSLLPSYMPISNI 2463
Db 1967 -----LPAYMVLRCU 1976
QY 2464 VLDKMLNANGVDRKELSRRAKVPVQKQTAAPLPTFPPISEVEVILCEATEVFGMKVD 2523
Db 1977 SKWTEPLTNGKIDRRAL----- 1995
QY 2524 ITDHFNLGHSLLATKLISRIDQRLKVRITVKDFDHPFADLASVIROGLGLOQPVSD 2583
Db 1996 -----PIPD 1999
QY 2584 GQODRSAHMAPRTETAILCDEFKVL-GFQVGITDNFFDLGGHSLMATKYLAVRIGRL 2642
Db 2000 ANVSRGVSYVAPRNGTEQKVADIWAQVLAQEQVAGYDHFEDIGHSLSAGMKMLALVHQEL 2059
QY 2643 DTTVSVKDVEDHPVLVQLALATLALDNLVQSKTNEIVGGREMAFYSFPQLLTFTDPEEFWASE 2702
Db 2060 GVLSLKDLPQSPTVEGLA-----QVIASAEKG-----TAAS 2091
QY 2703 IKPQLELQEIQDIYPTSQMKAF-----LFDHTTARPRFPVFPYIDFPSTSEPDAAAGLI 2757
Db 2092 ISPAEK-----QDTPVSSPQKRMVYLQOLEDAGTSYNMPAV-----LRLTGELDVERLN 2141
QY 2758 KACESLVNHLDIPTVFAEASGELYQVVLSCLOLPIQVITETDNINATNEFLDEFAKEP 2817
Db 2142 SVMQQLMORHEALRTTFEIKDGETVQRIWEEACEIAYFEAPEE---ETERIVSEFIK-P 2197
QY 2818 VRLGH-PLIFETLIKOTKSMRVM-RISHALYDGLSLEHVVRKLMHLYNGRSLPPH-QF 2874
Db 2198 FKIDQLPLFRIGLIKHSDEHVLDFMHHIISGASGVLEELSKYLDGETLEPLRIQY 2257
QY 2875 SRY-----MQYTAGRESGHGFV-RDVIQNTPMTILSD-----TVVDGNDATCKAL 2920
Db 2258 KDVAVMQOQFIQSELYKKQEEH--WLKELDGLPVPVTLPTDYSRPAVQTFEG-DRIAFSL 2314
QY 2921 HLSKIVNIPSOVLAR--GSSNIITQATVFNACALVLSRESKDVVGRIVSGRQGLPVE 2978
Db 2315 EAGK-----ADALRLRAKETDSTLYMVLLASYSAFLSKISQDDIIVGSPVAGRS--QAD 2367
QY 2979 YQDIVGPCTNAVPVRAHIES-----SDYNOLLHDIOQOYLLSLPHETIGFSDLRKNCNDWP 3034
Db 2368 VSRVIGMFVNTLALRYPKGKTFADY---LNEVKETALSADFAQDYPLEDLIGNVQVOR 2424
QY 3035 EAITN--FSCCITYHN-----FEYHPESQEQQRVEMGVLTKEFVNEMDEPLY 3080
Db 2425 DTSRNLPLDAVFSQONANIKDLTMKGIOLEPH---PERKTAK-----F 2465
QY 3081 DLAIAGEVEPDGAGLKVTVIAKTOLFCORRVE-----HLLEVS 3119
Db 2466 DLTLTAD-ETDG-GLTFVLEYNTALFKQETIERMKQYMMELLDVAVT 2509

RESULT 12
PSSL_BACSU
ID PSSL_BACSU STANDARD; PRT: 2561 AA.
AC P39845;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptide synthetase 1.
GN PFSA OR PPS1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95227362; PubMed=7711903;
 RA Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,
 RA Grandi G.,
 RT "A putative new peptide synthase operon in *Bacillus subtilis*: partial
 RT characterization.";
 RL Microbiology 141:645-648(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Katamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lepidus A., Lardinols S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 CC -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE
 CC (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z34883; CAA84360.1; .
 DR EMBL; Z99113; CAB13717.1; .
 DR HSSP; P14687; 1AMU.
 DR Subtilist; BG10970; ppsa.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR003880; Ppantne_attach.
 DR Pfam; PF00501; AMP-binding; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR Pfam; PF00668; Condensation; 3.
 DR PRINTS; PR00154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
 DR PROSITE; PS00455; AMP_BINDING; 2.
 DR PROSITE; PS00075; ACP_DOMAIN; 2.
 KW Multifunctional enzyme; Ligase; Repeat; Phosphopantetheine;
 KW Complete proteome.

| | | | | |
|----|----------|------|------------|----------------------------------|
| FT | DOMAIN | 966 | 1033 | ACYL CARRIER (ACP) 1. |
| FT | DOMAIN | 2012 | 2078 | ACYL CARRIER (ACP) 2. |
| FT | BINDING | 2042 | 2042 | PHOSPHOPANTHETHEINE (POTENTIAL). |
| SQ | SEQUENCE | 2561 | RA; 289180 | MW; 5476CBE4DD882FD2 CRC64; |

Query Match 9.9%; Score 1604; DB 1; Length 2561;
 Best Local Similarity 24.9%; Pred. No. 1.7e-84;
 Matches 611; Conservative 437; Mismatches 970; Indels 436; Gaps 87;

| | | | | |
|----|-----|---|--|-----|
| QY | 169 | VMTFS--- | HALVDVTFQORVLSRVPAAY--KHEKDHRTPEPSSDATDTSQSVSVSVMS | 224 |
| | | | | |
| DB | 128 | VWLFKAFHHIIMDGISLVNMGNOIIDLYQKMKKKDPLDPQPEPSYLSYTEKESQYLOS | 187 | |
| QY | 225 | CDNAVSAATHFW-QTHLNDLNASVFPHLSDHLMVNPNTTAA-----EHRITFPL | 272 | |
| | | | | |
| DB | 188 | FAKDRL---FWTQTFEHELE---YHSLADQTSLOKQSTASASRDITILSPDLEQTRIFC | 240 | |
| QY | 273 | SOKALSNSAICRTALSILLRSYTHSDEALFGA-VTEQSILPFDPKHYLDGTYQVAPLRVH | 331 | |
| | | | | |
| DB | 241 | EEHKINIITSLFMASFYICISRTISKDKLAIGTYVYGNRGSKAKEML--GMFVSSLPIT | 298 | |
| QY | 332 | CQS-----NLRASDVMDALISSYDDRLGLHPLAPGLDIRN | 365 | |
| | | | | |
| DB | 299 | VDPDDFLSVFTIGREQLSVMRHQRFPNLLVNLNLRNEQKDLNHLNIGLSMQYQLOWHN | 358 | |
| QY | 366 | TGDNGSAACDFQTVLLVTDGSHVNNNGINGFLQOITESSHFMPNCRNALLHCOMESGAL | 425 | |
| | | | | |
| DB | 359 | ADD-----FDYETALVFS--GYTANELSVQIQERID-----NGTI | 391 | |
| QY | 426 | LVAYDHNVIDSLQTRLLQOFGHLIKIQOS-----PLDLSMAE-VNLMTYDOR | 474 | |
| | | | | |
| DB | 392 | QLNFYQNTFLFSLDEIKRIQ--SHLLTILENAKHPHSFIRELDMTNTREKQKLYEFNK | 449 | |
| QY | 475 | AEIESWNSOPLEVQDTHLHHMLKAVSHSPTTAIOAWDGMWTSYELDNVSSRLAVHIKS | 534 | |
| | | | | |
| DB | 450 | TEAVSPKATL-----HGLFERQAFTPERIAIRFSGSLTYALDMMYASRLAAHLAA | 502 | |
| QY | 535 | LGRLAQOALIPVYFEKSKWVIASMLAVLKSNGNAFTLIDNPDPARTAAQVVTQTRATVALT | 594 | |
| | | | | |
| DB | 503 | RGV-TNESIVGLSERSPDLIAVLAVLAKGAGAYLPDPAYPKERLSYMLKDSGSLTLLT | 561 | |
| QY | 595 | SKLHRETOKLVGRCVVVDDELLOSVSASDDFSSLSKSDLAIVITFTSGTGPCKIMIE | 654 | |
| | | | | |
| DB | 562 | QP--GCCAPNFSGETLEVDMTSLAIEKAENHEFTPADGGSLAVIYVITSGTGPCKVAVE | 619 | |
| QY | 655 | HRAFSKALKFGASLGINSDFTRALQFGTHAFAGACILLEIMTTLINGCVCI--PSDDRMN | 712 | |
| | | | | |
| DB | 620 | HQAVSFLTGMHQHQPFLSEDDIVVMVKTSTSFDSVWQLFWWSLSGASAYLLPGWCKDSA | 679 | |
| QY | 713 | SIPSTFNRYNVNMMATPSYMGTFSP-----DYPGLATLVLVGEQSSSYNAITWA-- | 763 | |
| | | | | |
| DB | 680 | LIVQATHQENVTTAHFIPAMLSNFDQAEIERLSDRTSIKRVFAGGEPLAPRTAARFASV | 739 | |
| QY | 764 | -PKOLLNCGYGOSESSICFASNMSTEPNN-----MGRV-CAHSWVIDPDNDINRLVPI | 815 | |
| | | | | |
| DB | 740 | LPQVSLIHGYPTEATVDAAFYVLDPERDRDLRIPIGKVPGARIVLIDPH--LAVQPS | 797 | |
| QY | 816 | GAVGELVIESPGIARDYIVPPPEKSPFFTDTPSWYPANTFPDGAFLYRTGDLARASDG | 875 | |
| | | | | |
| DB | 798 | GVAGELYIAGAGVARGYLNRPALTEERFLED--PFYP-----GERMYKTGDVARMWLDG | 849 | |
| QY | 876 | STVCLGRIDSQVKIRGORVELGAIEFHLRQOQDDLTIVVEA--TKRSOSANSTSLIAFL | 933 | |
| | | | | |
| DB | 850 | NVEFLGRTDDQVKIRGYRIEPEGEIAALR-----SIEGVREAAVTVRTDS-GEPELCAV | 903 | |
| QY | 934 | IGSSYFGNRPSSDAHILDHDATKAINIKLEQVLPRHSIPSYICMLPELPTATKIDRRRL | 993 | |
| | | | | |
| DB | 904 | EG-----LQRNEVRA--QLQRLPLGYMVPAYMIEMEPMPTPSKGLDRNAL | 947 | |
| QY | 994 | RIMGKDILDKOTQGAIVQOAPAPIPVFADTAAKLHSIWVQSIGCIDPATNVGATPFELGG | 1053 | |
| | | | | |
| DB | 948 | PAPG-GAADAET-----YTAPRNV-----TEMKLSQWEDVLKNGP--VGIDHNPFRDGG | 994 | |

| | | | |
|----|------|--|------|
| Qy | 1054 | NSITAIKWVN - MARSVGMDLKVSINIOYHPITLAGISAVVK - GDPUSYTLIPKSTHSGPVEQ | 1111 |
| Db | 995 | HSLKATALVSIAKAEFEFQVPLKDDFAHPHTVEGLATVIRECTDSPYEAIPAEKQETYPV | 1054 |
| Qy | 1112 | SYSOGRLWFOLDOLGWSLWILIPYAVMRGPVNVDAARRALAAEQEHETILRTIFE - DQD | 1170 |
| Db | 1055 | SSAQRIYVLOOLEDGGTGYNMPAVLEGLKNPMERAFKELIKRHESLRTSFEODAG | 1114 |
| Qy | 1171 | GVGVOIVHEKLSBEMKVIDLCGSDLPFVYNQBOTT - -----PFLNSSE | 1214 |
| Db | 1115 | GDVQRHDEV - -----PFTL - ----QTIVLGERTEQEAFAAFIPKFDLSQA | 1155 |
| Qy | 1215 | AGWRATLLRLGEDDHILTIYMHHSIDGWSIDVLRDLNOLYSAAALKSDKDPJSAITPLP | 1274 |
| Db | 1156 | PLFRAQIVKISDERHLLLVMDHHTIISDGVSVNIIREFEGLYNN - -----RNLPALR | 1206 |
| Qy | 1275 | IQVSDFAKWOK - ----DOFTEQEKQNYWKQJLKDSPA - KIPTDFARPALLSGDAGCVH | 1328 |
| Db | 1207 | IQYKDYAVWREGFKTDAYKTQEA - --YWLKQLEGEPLFVLDPADHARPPVRFSFAGDKVS | 1263 |
| Qy | 1329 | VTIDGELYQSIRAFPCNHNHTTSFVLLIAAFRAAHYRLTAVEDAVIGTPPIANRNRPELEDI | 1388 |
| Db | 1264 | FTLDQEVASGLHKLARENGSTLWMLVAATAFSLRUSGQEDIIVGSPIAGRPHKDLPEI | 1323 |
| Qy | 1389 | IGCFVNTQCMRINIDHDHFTGTLINOVKATTTAAFENEDIPFERVVSALOPGSRDLSSTP | 1448 |
| Db | 1324 | LGMEVNTLALTRPEGGKPFQVQLQEVRETALEAFEHQDYFPEELVDKLEL - FDMXSRNP | 1382 |
| Qy | 1449 | LAQLIFAVHS - OKDLQRPFGQLESVPVPSKAYTR - ----FDMFHFUFQETSDLSKGVNF | 1502 |
| Db | 1383 | VFDAMFTLQVNEKQDI - ----DLREIKVRPANFAHSHISLFDITLATEISGSCCMBEF | 1436 |
| Qy | 1503 | ADELFKMETVENVVRVFETLRNGLOSSRTPVSTPLTDCIVITKLDVL - --NVKHVDY | 1559 |
| Db | 1437 | STEVFLKATIERWADHIFELHRAHS - --TPETSLAQINILSDKEKOKIVFEKNKQVEF | 1493 |
| Qy | 1560 | - PRESSLADYFQTVQSAYPDLSLAVDSSCLRTYTELDROQSDIAGWLRRMSPAETLVAV | 1618 |
| Db | 1494 | AQKDIPIPHRTFEAKAENPEHIAVIDNETEISYRLLNERANRLARTLQNRKGGKPT - VAV | 1552 |
| Qy | 1619 | FAPRSCETIVAFPGVLKANLAYLPDVRSPARVQDILSLSGPTIVLICHD - ----TAP | 1673 |
| Db | 1553 | LAKRSIDAIVGVILAVMKAGGVYIPDAHYPKARIEVILRD - SGADTLLQRELUKHLISNS - | 1611 |
| Qy | 1674 | PDEVTNVFVRIRDALNSDNADGFEVIEDHSTKPSATSLAYVLYTSGSTRGPKVYMIEH | 1733 |
| Db | 1612 | PESEMSHI - FLDDGGSFEESCNC - ----NLNLSPAPEEPVYIIYSGTTCAPKGVIVTY | 1663 |
| Qy | 1734 | RVIIRVTSGCIPNYP - -----SETRMAHMATIAFDGASYEIYSALLFGRTL | 1779 |
| Db | 1664 | Q - -----NFTHAALAWRQIVELDRKPVRLQIASFSDVFSGDILARTLTNGGTL | 1711 |
| Qy | 1780 | VCVDYMTTLDAKALKDVFFREHVNAASHVTSSSQDPLRVPRLRSRT - --LMFFFLVVTD | 1836 |
| Db | 1712 | IVCPDETRLEPABEYIKIISQRTIT - ----VMESTPALIIPVMEVYVRNQFKLPDLIDLILG | 1767 |
| Qy | 1837 | STAPDALDAOGLY - ----QGVOCYNGYGPTEGVYMSITIYPIDSTESFING - ----VPIGR | 1886 |
| Db | 1768 | SDMWKAQDKFTLDRFGQSMRIINSYGVTEATIDSSFY - ----ETSMGGECTGDNVPIGS | 1822 |
| Qy | 1887 | ALNNSGAYVDPBQQLVGIGVMGELYVTVDGLARGYSDK - ----ALDENRFRVHITVNDQ | 1940 |
| Db | 1823 | PLPNVHMVYLSQTDQIQPIGVAGELCIGGAGVAKGYHHRKPDLTQMKFTENPFV - ----S | 1876 |
| Qy | 1941 | TVKAYTGTDRVYRIGDGLIEFFGRMDTOFKIGNRIESAIEBAALLRROSSVDDAAVVLQ | 2000 |
| Db | 1877 | GERLYTGTDRACW - LPNGIRLLGRMDYQVYKNGYRIETEEIESVLLQTLGVREAAVAVQ | 1935 |
| Qy | 2001 | QNEQDQPEILGVVVAHDHSEN - ----DKGQSANOVEGWQDHFSGMYSDIGEIDPSTI | 2054 |
| Db | 1936 | HDKNQAGALYAIIVPS - DVNTNALRAALTKELPAYMIPAYLIPLVNMPLTNLNKLORNAL | 1994 |
| Qy | 2055 | GSPFKGWTSMYDGSQIDFDMHEWLGTEITRT - ----LHDNRSLGNVLEIGTGSMLFNLI | 2109 |

RESULT 13

ACVS_CEPAC

| ID | ACVS_CEPAC | STANDARD; | PRT; | 3712 AA; |
|----|------------|-----------|------|----------|
| | | | | |

AC P25464;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22; Last sequence update)

| DT | 16-OCT-2001 | (Rel: 40, Last annotation update) | (Rel: 40, Last annotation update) |
|----|-------------|-----------------------------------|-----------------------------------|
| DT | 16-OCT-2001 | (Rel: 40, Last annotation update) | (Rel: 40, Last annotation update) |

DE, Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase

DE (EC 6. - . -) (ACV synthetase) (ACVS).

GN PCBAB

OS Cephalosporium acremonium (Acremonium chrysogenum).

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina: Sordariomycetes:

Hypocreales; Hypocreaceae; mitosporic Hypocreaceae: Acremonium

OX NCBI_TaxID=5044;

RN [1]

RP SEQUENCE FROM N. A.

RX · MEDLINE=91177827: PubMed=1706706:
SEQUENCE FROM N.A.

RA MEDLINE-9117627, PubMed-1700700;
RA Gutierrez S., Diez B., Montenegro E., Martin J. F.:

"Characterization of the Cephalosporium acremonium nchar gene
Sutcliffe S.; Diez B.; Montenegro E.; Mallin J.F.;
NA
RT

RT encoding alpha-aminoacyl-tRNA synthetase a large

RT encoding alpha-aminoacyl-tRNA synthetase, a large multidomain peptide synthase: linkage to the rbcL gene as a cluster

RT-MULTIDOMAIN peptidase synthetase: Linkage to the *pcbc* gene as a clue

of early cephalosporin functional domains".

RT functional domains.";
T Bacteriophage 173-3254-3255-3256-3257-3258-3259-3260-3261-3262-3263-3264-3265-3266-3267-3268-3269-3270-3271-3272-3273-3274-3275-3276-3277-3278-3279-3280-3281-3282-3283-3284-3285-3286-3287-3288-3289-3290-3291-3292-3293-3294-3295-3296-3297-3298-3299-3300-3301-3302-3303-3304-3305-3306-3307-3308-3309-3310-3311-3312-3313-3314-3315-3316-3317-3318-3319-3320-3321-3322-3323-3324-3325-3326-3327-3328-3329-3330-3331-3332-3333-3334-3335-3336-3337-3338-3339-3340-3341-3342-3343-3344-3345-3346-3347-3348-3349-3350-3351-3352-3353-3354-3355-3356-3357-3358-3359-3360-3361-3362-3363-3364-3365-3366-3367-3368-3369-3370-3371-3372-3373-3374-3375-3376-3377-3378-3379-3380-3381-3382-3383-3384-3385-3386-3387-3388-3389-3390-3391-3392-3393-3394-3395-3396-3397-3398-3399-3400-3401-3402-3403-3404-3405-3406-3407-3408-3409-3410-3411-3412-3413-3414-3415-3416-3417-3418-3419-3420-3421-3422-3423-3424-3425-3426-3427-3428-3429-3430-3431-3432-3433-3434-3435-3436-3437-3438-3439-3440-3441-3442-3443-3444-3445-3446-3447-3448-3449-3450-3451-3452-3453-3454-3455-3456-3457-3458-3459-3460-3461-3462-3463-3464-3465-3466-3467-3468-3469-3470-3471-3472-3473-3474-3475-3476-3477-3478-3479-3480-3481-3482-3483-3484-3485-3486-3487-3488-3489-3490-3491-3492-3493-3494-3495-3496-3497-3498-3499-3500-3501-3502-3503-3504-3505-3506-3507-3508-3509-3510-3511-3512-3513-3514-3515-3516-3517-3518-3519-3520-3521-3522-3523-3524-3525-3526-3527-3528-3529-3530-3531-3532-3533-3534-3535-3536-3537-3538-3539-3540-3541-3542-3543-3544-3545-3546-3547-3548-3549-3550-3551-3552-3553-3554-3555-3556-3557-3558-3559-3560-3561-3562-3563-3564-3565-3566-3567-3568-3569-3570-3571-3572-3573-3574-3575-3576-3577-3578-3579-3580-3581-3582-3583-3584-3585-3586-3587-3588-3589-3590-3591-3592-3593-3594-3595-3596-3597-3598-3599-3600-3601-3602-3603-3604-3605-3606-3607-3608-3609-3610-3611-3612-3613-3614-3615-3616-3617-3618-3619-3620-3621-3622-3623-3624-3625-3626-3627-3628-3629-3630-3631-3632-3633-3634-3635-3636-3637-3638-3639-3640-3641-3642-3643-3644-3645-3646-3647-3648-3649-3650-3651-3652-3653-3654-3655-3656-3657-3658-3659-3660-3661-3662-3663-3664-3665-3666-3667-3668-3669-3670-3671-3672-3673-3674-3675-3676-3677-3678-3679-3680-3681-3682-3683-3684-3685-3686-3687-3688-3689-3690-3691-3692-3693-3694-3695-3696-3697-3698-3699-3700-3701-3702-3703-3704-3705-3706-3707-3708-3709-3710-3711-3712-3713-3714-3715-3716-3717-3718-3719-3720-3721-3722-3723-3724-3725-3726-3727-3728-3729-3730-3731-3732-3733-3734-3735-3736-3737-3738-3739-3740-3741-3742-3743-3744-3745-3746-3747-3748-3749-3750-3751-3752-3753-3754-3755-3756-3757-3758-3759-3760-3761-3762-3763-3764-3765-3766-3767-3768-3769-3770-3771-3772-3773-3774-3775-3776-3777-3778-3779-3780-3781-3782-3783-3784-3785-3786-3787-3788-3789-3790-3791-3792-3793-3794-3795-3796-3797-3798-3799-3800-3801-3802-3803-3804-3805-3806-3807-3808-3809-3810-3811-3812-3813-3814-3815-3816-3817-3818-3819-3820-3821-3822-3823-3824-3825-3826-3827-3828-3829-3830-3831-3832-3833-3834-3835-3836-3837-3838-3839-3840-3841-3842-3843-3844-3845-3846-3847-3848-3849-3850-3851-3852-3853-3854-3855-3856-3857-3858-3859-3860-3861-3862-3863-3864-3865-3866-3867-3868-3869-3870-3871-3872-3873-3874-3875-3876-3877-3878-3879-3880-3881-3882-3883-3884-3885-3886-3887-3888-3889-3890-3891-3892-3893-3894-3895-3896-3897-3898-3899-3900-3901-3902-3903-3904-3905-3906-3907-3908-3909-3910-3911-3912-3913-3914-3915-3916-3917-3918-3919-3920-3921-3922-3923-3924-3925-3926-3927-3928-3929-3930-3931-3932-3933-3934-3935-3936-3937-3938-3939-3940-3941-3942-3943-3944-3945-3946-3947-3948-3949-3950-3951-3952-3953-3954-3955-3956-3957-3958-3959-3960-3961-3962-3963-3964-3965-3966-3967-3968-3969-3970-3971-3972-3973-3974-3975-3976-3977-3978-3979-3980-3981-3982-3983-3984-3985-3986-3987-3988-3989-3990-3991-3992-3993-3994-3995-3996-3997-3998-3999-4000-4001-4002-4003-4004-4005-4006-4007-4008-4009-4010-4011-4012-4013-4014-4015-4016-4017-4018-4019-4020-4021-4022-4023-4024-4025-4026-4027-4028-4029-4030-4031-4032-4033-4034-4035-4036-4037-4038-4039-4040-4041-4042-4043-4044-4045-4046-4047-4048-4049-4050-4051-4052-4053-4054-4055-4056-4057-4058-4059-4060-4061-4062-4063-4064-4065-4066-4067-4068-40

RL J. Bao
PM 137

RN [2]

RP . PARTIAL SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=ATCC 11550;

RX MEDLINE=91168300; PubMed=20765552;

RA Hoskins J.A., O'Callaghan

RA Chen V.J., Skatrud P.L.;

RT "Gene disruption of the pcba

RT Cephalosporium acremonium.";

RL Curr. Genet. 18:523-530(1990).

CC - I - FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE

CC . ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH-PEPTIDE BOND

CC FORMED THROUGH

CC INTERMEDIATES.

CC - i - COFACTOR: CONT

CONCENTRATION, CONTAINING 3 COEFFICIENTS FOUND FROM ROTARY PLATES
(POTENTIAL).

CC -I- PATHWAY: FIRST (CONTINUED).

CC 1. SUMMARY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
CC CEPHALOSPORIN.

CC - ! - SIMILARITY: BE

CC ; SIMILARITY, BELONGS TO THE AIR DEFENDANT AM
CC FAMILY.

CC, -I- STIMILARITY: CONT

Db 2417 GVLNPEP---TNLOPSRDSGNSLHGLFEDIVASTPDRIAIACTRLSLSELNERANQL 2473
 Qy 1601 LAGWLRRSRMPAETLVAVFAPRSCETIVAFGVLKANLAYLPLDVRSPSARVQDILSGLS 1660
 Db 2474 VHLIISASIVADDRIALLDKSIDMVIALLAVKAGAAVPLDPTPSQRTLEILESS. 2533
 Qy 1661 GPTIVLGHOTAPPDIEVTWVEFVRI--RDALNDSNADGVEVIEHDSKPSATSLAVLY 1718
 Db 2534 ARTLITTRKHT-PRGGTVANVPVSVLDSPETLACLNQSKENPTTSTQKPS--DLAYVIF 2590
 Qy 1719 TSGTGRPKGMIEHVIITVTSGCIPNYPSETRMAH---MATIAFDGASVEIYSALL 1774
 Db 2591 TSGTGRPKGVLEHQSVQV--LNSLIERFGETNGSHAVLFUSVYDFSLQOLCUSVL 2649
 Qy 1775 FGRTLVCVDMYMTILDARALKDVFREHVHNAASHVTSQQDVL-VPVRLSRITLMFFFLV 1833
 Db 2650 GGNKLI-IPPEGLTHEAFYDGRREKLSVLSGTPSVLQIELSRLPH----LHMVTA 2703
 Qy 1834 VTDSAPDALDAQLOGVOCYNGYGTENGVMSTIYPIDTSFNGVPIGRAL-----N 1889
 Db 2704 GEEFHASQFEKMRSQFAG-QINNAYGITE---TTVYNIITT--FKGDAPFTKALCHGIP 2756
 Qy 1890 NSGAYVDPPEQQLVGIGVMGELVVTGDLAGY-SDKALDENRFV-----HITVNDOT 1941
 Db 2757 GSHVYVNDLRLQVFPFNAVGEVLYGGDCLARGYLNQDALTNERFIPNPFYEPKQASDRP 2816
 Qy 1942 VKAYRTGDRVRYRIGDGLIEFFGRMDQFKIRGNRIESAETEAALLRDSVSRDAVVLQ 2001
 Db 2817 QRLYKTKDGLVRFER-GPHILEYLRGKQKQVLRGFRIELSEVRDAVLAISAVKEAAVPIKY 2875
 Qy 2002 NED 2004
 Db 2876 DED 2878

RESULT 14

ACVS_NOCLA ACVS_NOCLA STANDARD; PRT; 3649 AA.
 ID ACVS_NOCLA
 AC P27743
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Delta (L-alpha-aminocadipyl)-L-cysteiny-D-valine synthetase
 GN (EC 6.---) (ACV synthetase) (ACVS).
 OS PCAB.
 OS Nocardia lactandurans.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
 OC Amycolatopsis.
 OC NCBI_TaxID=1913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VAR LC 411;
 RX MEDLINE=92065808; PubMed=1956290;
 RA Coque J.J.R., Martin J.F., Calzada J.G., Liras P.;
 RT "The cephamycin biosynthetic genes pcAB, encoding a large
 RT multidomain peptide synthetase, and pcBC of Nocardia lactandurans are
 RT clustered together in an organization different from the same genes
 RT in *Acromonium chrysogenum* and *Penicillium chrysogenum*.";
 RL Mol. Microbiol. 5:1125-1133(1991).
 CC -1- FUNCTION: EACH OF THE CONSTITUTED AMINO ACIDS OF THE TRIPEPTIDE
 CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS
 CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
 CC INTERMEDIATES.
 CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
 CC CEPHALOSPORIN.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X57310; CAA040561.1; -
 DR PIR; S18268; S18268.
 DR HSP; P14687; IAMU.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR003880; Ppancne_attach.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF005501; AMP-binding; 3.
 DR Pfam; PF00658; Condensation; 3.
 DR Pfam; PF00975; Thioesterase; 1.
 DR PRINTS; PR00154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
 DR PROSITE; PS00455; AMP_BINDING; 1.
 DR PROSITE; PS00075; ACV_DOMAIN; 3.
 DR Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
 KW Repeat; Phosphopantetheine.
 KW REPEAT 401 861 DOMAIN 1 (ADIPATE-ACTIVATING);
 FT REPEAT 1014 1937 DOMAIN 2 (CYSTEINE-ACTIVATING).
 FT REPEAT 2079 2985 DOMAIN 3 (VALINE-ACTIVATING).
 FT DOMAIN 788 857 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1864 1933 ACYL CARRIER (ACP) 2.
 FT DOMAIN 2910 2981 ACYL CARRIER (ACP) 3.
 FT BINDING 820 820 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 1896 1896 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 2944 2944 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT ACT_SITE 3502 3502 THIOESTERASE (BY SIMILARITY).
 SQ SEQUENCE 3649 AA; 404079 MW; 6FD095704F858E6B CRC64;

Query Match 9.2%; Score 1484.5; DB 1; Length 3649;
 Best Local Similarity 21.8%; Pred. No. 2.7e-77;
 Matches 686; Conservative 424; Mismatches 1071; Indels 963; Gaps 104;
 Qy 90 AKWEIVNQTALR-AFAFTSDSGKTSQVILKDSFVSMCWSSSSDPVVRDEAAAA-- 146
 Db 923 AWQAARQTYPALRLRFDWAE--PVOIVNDNDKPFWRFDVLSATADDAEQARVRELQ 979
 Qy 147 -----ASGPRCNRVLL---EDMTKKCOLVWTFSHALVDVTFOQVLSRVFAAYK 194
 Db 980 ERDTEPYDLAGGLRFLVLIKQEDL---FSLFSCHHILDGWLSPLVLDHDEVHRNYL 1035
 Qy 195 HEKDTHRPETESSDATDTSQSVSVSMSCEDNA-VSATHFWQTHLND----- 242
 Db 1036 -----ALRGAQPIESDV-----DNAYVAAQRYWEAHRNDHAAYVWEOLGR 1075
 Qy 243 -----LNASVFPPLS-----DHLVNPVNTTAEHR-----ITFPLSQKALS 279
 Db 1076 IDERGFAGLLNEKSRVRSVLDGYDHVQ-----RHRTKLYLGADLIGALKAGCAAD 1127
 Qy 280 SAICRTALSTLLSRYTHS-----DEALFGA-VTQSLPFDKHYLDGTYQTVLRVHCQS 334
 Db 1128 QVTLHSLVQFVWHKVLHAIGGNTTVGTIVSGRNLPGVDIGENSAGLFIINTLPL----- 1181
 Qy 335 NLRASDVMDAISYDDRLGLHAPFGLRDIR---NTGDNGSAA-----CDFQTV 379
 Db 1182 -----IVDDQAGQONVAEVRDIOAVNTMNSKISIVELGLRQSGEMKRRLEDTL 1231
 Qy 380 LLVTGDSHVNNNGINGFLQQTITSSHPMPCNNRAI-----LHCOMESSGAL 425
 Db 1232 LVLENYPRL-----LDEEELAH-----QEARFEKAYDAKVDYPIAVVAREEGDEL 1279
 Qy 426 LVA-----YYDHNVIDS-LQTTRLLOQFGLHKLQIQLSPDLDSMAEVNLMTEYDRAEIE 478
 Db 1280 TVTLWYAGELFDEDTITDLVDARTL--FRQVTEIDIARP-----VRELDLISPMRARD 1332
 Qy 479 SWNSQPLEV-QDTLIIHHEMLKAVSHSPTKTAIQAWGDWTYSLELDNVSSRLAVHIKSLGL 537

[illegible]


```
QY 2579 QPVSDGQQRSAHMAPRTTEAILCDEFKVLGFQ-VGITDNFFDLGGHSLMATKLAVR 2637
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3005 -----GNPQHQISYNPPRDLVLEADLCRLWASALGTERCGIDDDDLFRLGDSITALTHAAQ 3059
QY 2638 IGHRLDTTVSKOVDFDHPVLFQALALDNLV-----OSKTNEIYGGREMAEYSP 2686
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3060 IHQIGRKVTVRDIFDHTTIRGHIH---DNVWVKLVPHNVFQFQAEQQTVLGD---APLLP 3113
QY 2687 FQLLTEDPEEFMASEIKPOLELQELIIDIYPSTQMOKAFLEFDTHTTARPRPFVFFYIDFP 2746
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3114 IQIWFLE-----SKSLQHPSHWNT-----FYL--- 3135
QY 2747 STSEPDAAGLIKACESLVNHLDFIRTVFAEASGELYQVVLSCLDLPIQ--VIETED---- 2800
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3136 RTPDLDTTTLSTAVAEQLYHDAFRMLRQIDGRTVQCFADDIS-PVQLRVLVNKVDGDS 3194
QY 2801 -NINTATNEPDEF--AKEPVRGLGHPLEIRFTIITKQKSMRVIMRISHALYDGLSLEHVVR 2857
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3195 AADQQLQKQSDFDLEKGPICAAAYLHGZ---EDRSARWFVSVHHIIIDIYSWQLLAR 3250
QY 2858 KLHMLYN-----GRSLLPPHOFSRYMQYADGREGHGFWRDVIQNT-----PMT--- 2902
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3251 DLQILYEGGTGRKSSSVRWAEALQ-SYQGSASERAYWEGLLAQTAANISALPPVVTGTR 3309
QY 2903 -----ILSDDTVV-----DGNATCKALHLSKIVNIPSOVLRGSSNIIT 2941
| | : : | | : : | | : : | | : : | | : : | | : :
Db 3310 TRLARTWSDDRVTILLNEASNQASTODLLLAAGVGLAQOQVTPGSPSMT 3359
```

Search completed: May 30, 2003, 12:46:02
Job time : 139 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2003, 12:40:46 ; Search time 78 Seconds
(without alignments)
3856.469 Million cell updates/sec

Title: US-09-482-788-2
Perfect score: 16128
Sequence: 1 MEYLTAVDGRDLPPTPASF.....RVEHLLLEVSKEFEGNSSL 3129
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----------|---|
| 1 | 9720 | 60.3 | 3131 | 2 S39842 | enniatin synthetase |
| 2 | 5000 | 31.0 | 15281 | 2 S41309 | cyclosporin synthetase |
| 3 | 2275.5 | 14.1 | 4848 | 2 T30289 | pristinamycin I synthetase |
| 4 | 2084 | 12.9 | 6486 | 2 T31076 | tyrocidine synthetase |
| 5 | 1984 | 12.3 | 1997 | 2 T30874 | virginiamycin S synthetase |
| 6 | 1975 | 12.2 | 6359 | 2 T31679 | bacitracin synthetase |
| 7 | 1933 | 12.0 | 4450 | 2 JX0340 | gramicidin S synthetase |
| 8 | 1906.5 | 11.8 | 3670 | 2 T36249 | CDA peptide synthetase |
| 9 | 1896.5 | 11.8 | 2588 | 2 AI2136 | peptide synthetase |
| 10 | 1882.5 | 11.7 | 5255 | 2 T31677 | bacitracin synthetase |
| 11 | 1876.5 | 11.6 | 9376 | 2 T14593 | syringomycin synthetase |
| 12 | 1867.5 | 11.6 | 5149 | 2 F83345 | probable non-ribosomal peptide synthetase |
| 13 | 1816 | 11.3 | 3587 | 2 T31075 | tyrocidine synthetase |
| 14 | 1810 | 11.2 | 7463 | 2 T36248 | CDA peptide synthetase |
| 15 | 1807.5 | 11.2 | 2555 | 2 C69681 | peptide synthetase |
| 16 | 1793.5 | 11.1 | 3587 | 2 I40486 | surfactin synthetase |
| 17 | 1781.5 | 11.0 | 2617 | 2 AE2136 | peptide synthetase |
| 18 | 1769.5 | 11.0 | 2560 | 1 I40457 | peptide synthetase |
| 19 | 1758 | 10.9 | 4452 | 1 YG8SG2 | gramicidin S synthetase |
| 20 | 1729.5 | 10.7 | 5369 | 2 T44807 | mycosubtilin synthetase |
| 21 | 1725 | 10.7 | 3603 | 1 D69681 | peptide synthetase |
| 22 | 1710 | 10.6 | 2607 | 2 T31678 | bacitracin synthetase |
| 23 | 1708.5 | 10.6 | 2448 | 2 S53999 | pyoverdine synthetase |
| 24 | 1677 | 10.4 | 4342 | 2 H83343 | probable non-ribosomal peptide synthetase |
| 25 | 1672 | 10.4 | 2611 | 2 T14591 | actinomycin synthetase |
| 26 | 1625.5 | 10.1 | 5232 | 2 A45086 | HC-toxin synthetase |
| 27 | 1612.5 | 10.0 | 2566 | 2 E98274 | hypothetical protein |
| 28 | 1611.5 | 10.0 | 2117 | 2 T36180 | CDA peptide synthetase |
| 29 | 1604 | 9.9 | 2561 | 1 I40456 | peptide synthetase |

| | | | | | |
|----|--------|-----|-------|----------|---|
| 30 | 1601 | 9.9 | 3588 | 2 I40485 | surfactin synthetase |
| 31 | 1538.5 | 9.5 | 4976 | 2 T14165 | peptide synthetase |
| 32 | 1532 | 9.5 | 2157 | 2 AI3009 | peptide synthetase |
| 33 | 1515 | 9.4 | 3712 | 1 YGCEVC | alpha-aminoadipyl-pristinamycin I synthetase |
| 34 | 1491 | 9.2 | 2591 | 2 T30288 | pristinamycin I synthetase |
| 35 | 1484.5 | 9.2 | 3649 | 1 S18268 | delta-(L-alpha-aminoadipyl)-enniatin synthetase |
| 36 | 1453.5 | 9.0 | 492 | 2 S53111 | delta-(L-alpha-aminoadipyl)-enniatin synthetase |
| 37 | 1422.5 | 8.8 | 3770 | 2 A40889 | hypothetical protein |
| 38 | 1402 | 8.7 | 2391 | 2 G89779 | probable non-ribosomal peptide synthetase |
| 39 | 1391.5 | 8.6 | 10797 | 2 T30192 | saframycin Mx1 synthetase |
| 40 | 1387 | 8.6 | 1317 | 2 B83346 | hypothetical protein |
| 41 | 1366.5 | 8.5 | 2605 | 2 T18552 | alpha-aminoadipyl-lysobactin synthetase |
| 42 | 1358 | 8.4 | 4077 | 2 T17484 | hypothetical protein |
| 43 | 1347.5 | 8.4 | 3791 | 1 YGPIV8 | alpha-aminoadipyl-lysobactin synthetase |
| 44 | 1346.5 | 8.3 | 3746 | 1 YGPIV3 | lysobactin synthetase |
| 45 | 1331 | 8.3 | 1575 | 2 T18545 | lysobactin synthetase |

ALIGNMENTS

RESULT 1

S39842
enniatin synthetase - fungus (Fusarium scirpi)
C:Species: Fusarium scirpi
C>Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 03-Nov-2000
C:Accession: S39842; S35906; S65363
R:Haese, A.
submitted to the EMBL Data Library, November 1992
A:Reference number: S39842
A:Accession: S39842
A:Molecule type: DNA
A:Residues: 1-3131 <HA>
A:Cross-references: EMBL:Z18755; NID:g2729; PIDN:CAA79245.1; PID:g2730.
R:Haese, A.; Schubert, M.; Herrmann, M.; Zocher, R.
Mol. Microbiol. 7, 905-914, 1993
A:Title: Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme enniatin synthetase
A:Reference number: S35906; MUID:93247491; PMID:8483420
A:Accession: S35906
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 499-1074;1572-1988;2423-2566 <HA2>
A:Cross-references: EMBL:Z18755
A:Experimental source: strain ETH 1536/J5
R:Pieper, R.; Haese, A.; Schroeder, W.; Zocher, R.
Eur. J. Biochem. 230, 119-126, 1995
A:Title: Arrangement of catalytic sites in the multifunctional enzyme enniatin synthetase
A:Reference number: S65363; MUID:95324513; PMID:7601090
A:Accession: S65363
A:Molecule type: protein
A:Residues: 2029-2048;430-437;1011-1020;1021-1034;1677-1695;2294-2299 <PIE>
A:Experimental source: strain ETH 1536/J5
C:Genetics:
A:Gene: esyn1
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein; multifunctional enzyme; phosphopantetheine; phosphoprotease
F:531-985/Domain: acetate-CoA ligase homology <ACLI1>
F:1603-2100/Domain: acetate-CoA ligase homology <ACLI2>
F:2507-2574/Domain: acyl carrier protein homology <ACP1>
F:2601-2667/Domain: acyl carrier protein homology <ACP2>
F:1047,2538,2632/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 60.3%; Score 9720; DB 2; Length 3131;
Best Local Similarity 60.8%; Pred. No. 0;
Matches 1939; Conservative 443; Mismatches 691; Indels 114; Gaps 32;
QY 1 MEYLTAVDGRDLPPTPASFCSHGSDPLNSSYQQLFHLVYGLDSRIEAIKPTFPFQDLM 60
Db 1 MSLHTPSDGGQDPALASKTLK-----EQISRALGLGQDKIENIPGTFPFQDVI 49
QY 61 DCNALDKOSAGHAYVDVPTDIDISRFALAWKEIVNOTPALRAFAFTSDSGSKTSOVLTKD 120
Db 50 DCAADDKQRAVGHAVFEIPKIDIDARLAAAWKETVLHTPALTRCTCTTYSKSGDVLQVVLRD 109

| | | | | | |
|----|-------|---|----|-------|--|
| QY | 923 | -----SANSTSLIAFLIGSSYFNGRPSDAHILHDHATKA-----INIKLEOVL 965 | Db | 13057 | YQIEFFGRLDQOIKIRGHRIEPAEVEQALLSDSSINDAVVSYAQNK-EGLEMGVITTOA 13115 |
| Db | 12029 | PELVAFPSLKNANGTNGV-----NGVSDOEKIDQEQHALLMENKIRINLQALL 12078 | QY | 2018 | DHSENDKOSANOVQEGQDHEESGMYSDIGEDIDSTIGSDFKGTSMYDGSQIDDEHME 2077 |
| QY | 966 | PRHSIPSYICMLPELPTATGKIDRRRLRIMGKIDLDKQTOCAIVQOAPAPIPVADTAA 1025 | Db | 13116 | AQSV-DKEEASNKVQWEAHEFDSTAYANIGGIDRDALQGDQFLSTSMYDGLIFREEMQE 13174 |
| Db | 12079 | PTYMPSKRIHVDQLPVNANGKIDRNELAVR-----AQATPTSSYSTYVAP 12125 | QY | 2078 | WLGTTTRTLHDNRSLGNVLEIGTSGMILFNL--SRLESYVGLPEPSRASAFAFNKAPES 2135 |
| QY | 1026 | K--LHSTWQSLGIDPATVNVGAT--FFEGLGNSITAIRM-VNMAVSQMDLKYNSIYQH 1080 | Db | 13175 | WLNDTMSLLDNQPPGKVLIEGTGVMVFNGLKVEGLOSAGLEPSRSVTAWNKALET 13234 |
| Db | 12126 | RNDIETIICKEFA--DILSVRVGIIDNFDLGGHSLIATKLAARLSRRLDTRVSRVDVDT 12184 | QY | 2136 | IPSLAGAKVQVGTATDIGQVDDLHDPDLVNLVNSVQYFPPSEYLAETADTLHLPLNVQRI 2195 |
| QY | 1081 | PTLAGISAVVKGDPLSTLTPKSTHEGPEQSYSGRLWELQDLQDGLSLWYLIPIAYMR 1140 | Db | 13235 | FPSLAGSARVHVGTAEIDISSIDGLRSLVINSVAQYFPPSREYLAELTANLIRLPGVKRI 13294 |
| Db | 12185 | PVQCILAASQOGSTPEALPULSHSGPVQOOSFAQGRLEWELFELDNAAVIMPFVGLR 12244 | QY | 2196 | FFGQVRSOATNEHEFLAARAIHTLGKKNATKDDVROKMAELEDEMEELLVEPAFFTSKDRF 2255 |
| QY | 1141 | GPVNVDALRALAALEORHETLRTTFEDQGVGVQIVHEKLSEEMKVIDLCGSLDPEFV 1200 | Db | 13295 | FFGDMRYATNKDFLARAVHTLGSNASKAMVROQVAKLEDEBELVDPAFFTSLSQOF 13354 |
| Db | 12245 | GPLRVDALQALRALEERHELLRTTFEEDQGVGMQIVHSPMRDICCVDISGANED-LAK 12303 | QY | 2256 | PGLVEHVEIILLKNNMAYNELSARYAAVHVHGRSL--GDE-----LVLPEKDDWDIFQ 2307 |
| QY | 1201 | LNQEOITPFFNLSSPAGHRATLLRLGEDDHILTIVMHHIISDGSWIDVLRDLNOLYSAAL 1260 | Db | 13355 | PDEIKHVEILLPKRMAATNELSSRYAAVHVGHGHPNGEDEDQWAVKDINPKAWDFA 13414 |
| Db | 12304 | LKEEQQAPFNLSTEVAVRWALFKAGENHHILSIIVMHHIISDGSWIDVIFQOELAQFYSAV 12363 | QY | 2308 | ANOLNOKSLGDLKSSDAA--IMAVSKIPEITAFERQOVVASLNSNID-----EWOL 2357 |
| QY | 1261 | KDSKDPLSALTPLPIQYSDFAKWOD--QFIEQEKOLNYKKOLKSSPAKIPTDFARPA 1318 | Db | 13415 | GTRMDRQALLQLODRQGDVVAVSNIPYSKTIIMERHLSQSLDDDDGTSAVDGTAW-I 13473 |
| Db | 12364 | R-GHDPLOSQVKPLPIHYRDFAVWQROKQVAVHESQLOYWIEQLADSTPAEILSDFNRPE 12422 | QY | 2358 | STIRSSAEGSSSLSPDIFRIAGEAGFERVESSAROWSQNGALDAVHF-----HCCSQG 2411 |
| QY | 1319 | LLSGDAGCVHVTIDGELYQSLRFAFCNEHNTTSFVLLAARAAHYRLTAVEDAVIGTPIA 1378 | Db | 13474 | SRTQSRAKECPALSVADLIEIGKIGIGEVESAWARQHSQGLDADAFHREPPRH--SS 13530 |
| Db | 12423 | VLSGEACTVPIVEDEYEXLSLFCRNHQVTSFVLLAARFAHYRLTGAEDATIGTPIA 12482 | QY | 2412 | RTLVPPTDHLHKGSLLTNRPLQRQNRRIAEVRRLRLSLPSYMIPIVVLDMPL 2471 |
| QY | 1379 | NRRPELEDLIGCFVNTQCHIRINIDHDTGLTINQVKATTTAFENEDIPFERVSALQ 1438 | Db | 13531 | HVMFERPTEKHGRSSSLTNRLPLHLLQSRLEAKVRERLQSLPPYMIPIKRTILDDQML 13590 |
| Db | 12483 | NRRPELEDLIGCFVNTQCHIRIALEEHDFNSVVRVRSFAAFENQDVPFERLVSALL 12542 | QY | 2472 | NANGKVDKELSRKVPKQOATAAPLTPFISEVEVILCEATEVECMKVIDTIDHFNFL 2531 |
| QY | 1439 | PGSRDLSSTPLAQILFAVHOSQOKLGRFKFQGLSPVPSPKAYTRFDMERHFLFOETSLKG 1498 | Db | 13591 | TSNGKVDKRLAKQARVPIPSAASTLDFVAPRTEIEVVLCEETDGLLVKVGITDNFEL 13650 |
| Db | 12543 | PGSRDASRNPLVQLMFVHVSQORNKLGKLEGEPTPYTATTFDFVEHFLFQDKGLAG 12602 | QY | 2532 | GHSLLATKLISRIDQRLKVRITVKQVDFHPFADLASVIRQG-----2574 |
| QY | 1499 | SVNFADLPMKTEVNVVRVFFELRLNGLOSSRTPVSIPLTDGIVLEKLDVNLVKHV- 1557 | Db | 13651 | GHSLLATKLARSRLRLDAGITVKQVFDQPLADLAASILQSSRRHSRIPSUPYEGPVE 13710 |
| Db | 12603 | NVYFAADLFAAATIRSVVEVFEHLRGLRQDQDIAISTMPLVDGLAALNSRLPAVEDIE 12662 | QY | 2575 | -----2574 |
| QY | 1558 | -DYPRESSLADVFQTOVSAYPDSLAVVDSSCLTYTELDRQSDILAGWRRRRSMPTLV 1616 | Db | 13711 | QSPAQGLWFLDQFNIDALWYLIPFALMRGPLOQVDALAAALVALEERHESLRTTFEERD 13770 |
| Db | 12663 | PDFAEASVVDVQTOVAMPDALAVTDTSTKUTYAELOQSDSDHVAAMLSKQKLPAESIV 12722 | QY | 2575 | -----2574 |
| QY | 1617 | AVFAPRSCETIVAFFGKANLAYPLDVRSPSARVQDILSLGSGPTIVLIGHDTAPPDI 1676 | Db | 13771 | GVGIQVVPPLRTTKDIRIDVSGNRDDADYLEPLQKEQOQTPDLASEPGWRVALLKLGKD 13830 |
| Db | 12723 | VLAPRSSETIVACIGILKANLAYLPMSNVPEARQAILSEIPGEKFLVLLGAGVIPDN 12782 | QY | 2575 | -----2574 |
| QY | 1677 | EVTNVFEVRIRDALNSNADGFVIEHSDTKPSATSLAVLYTSGTGRPKGVMIHRVI 1736 | Db | 13831 | DHILSIVMHHIISDGNSTEVQLRELQFYLAAKSGKAPLSQVAPLPIQYRDFAVWQROBE 13890 |
| Db | 12783 | KTADVRMVFISDIAVASKTDKYS----PGTRPSASSLAYIETSGTGRPKGVMEHGV 12838 | QY | 2575 | -----2574 |
| QY | 1737 | IRVTSGCIPNVPSETPRMAHMAFIADGASYEYISALLFGRTLVCVYDWTIDARALKDV 1796 | Db | 13891 | QVAESORQLDYKKQLADSSPAELLADYTRPNVLSGEAGSVFVINDSVYKLSVFCRSR 13950 |
| Db | 12839 | ISLVKQNA-SRIPQSLRMAHVSMLAFDASWELFTLLNGGLTFCISYFTVLDLSKALSA 12897 | QY | 2575 | -----2574 |
| QY | 1797 | FFREHVNAAHSVTSQSDQVPLRVRPRLSRTLMPFFLVVTDSTAPDALDA-QGLYQG- 1851 | Db | 13951 | QVITFTTLLAAPRAAHYRWGSDDATIGTPIANRNPPELENLIGCFVNTQCMRITIGDDE 14010 |
| Db | 12898 | FSDHRIN-----ITLLPALLKQCL-----ADAPSVLSSLESILYIGDRL 12937 | QY | 2575 | -----2574 |
| QY | 1852 | -----VQCYNGYGPTEGVMSITIPIIDSTESFINGCVPTGRALN-NSGAYVVD 1897 | Db | 14011 | TFESLVQVQVRSTATAFENQDVPERIVSTLSAGSRDTSRNPLVQLLFAVHSQOGLRIQ 14070 |
| Db | 12938 | DGADATKVDKLVKGKAYNAIGTENSVMSTIYIEH-ETFANGVPIGTSLGPKSKAYIMD 12996 | QY | 2575 | -----LG 2576 |
| QY | 1898 | PEOQLYIGVMGELVVTGDLARGYSDKALDENRFRVHTVNDQTKAYRTGDRVRYRIGD 1957 | Db | 14071 | LCGVDPVLSVTSTFRDLEFHAFQPADRLNGSVMFATDLFQPETIQGFVAVVEVLQRG 14130 |
| Db | 12997 | QDOQLYPAGVMGELVAGDLARGYTDPSLNTGRFIHTIDGKQVQYRTGDRVRYRPRD 13056 | QY | 2577 | LOQPV- 2582 |
| QY | 1958 | GLIEFFGRMDTOFKIRGNRIESAEIEBAALLRDSVRDAAVVLQONEDQAPEILGFVYVADH 2017 | | | |

Db 450 QV-----LEENDTGLPAEDATLAALPEAQARTPDPTTALLVGRSLTYAELNARNR 502
QY 528 LAVHKSGLRAOQAIIPIYFEKSKWIASMLAVLKSNAFTLIDNDPDPARTAQVVTOT 587
Db 503 LARHLVTLGAGPQ-IVAVKLESLDYVALLAVLKTGAAYLPVDYTAERAFWMDDA 561
QY 588 RATVALTSKLRHTVOKLVGRCVVDDQLQSVASDDFSSLTUKSODL-----AYVIFTS 642
Db 562 RPAVLTGE---DTGQDLSC---YDD-----TDLTDADRTAPLLPAHPAYVYTS 605
QY 643 GSTGDPKGMIEHRAFSKALKEGASIGINDSTRALQFGTHAFGACLELIMTTLINGCV 702
Db 606 GSTGTAKVMPGGAALVLLNHRREITPGEAGAPVAQFTTIGDVAAQEILATWHLGKTL 665
QY 703 CIPSDDDRMNS---IPSFIRINRYN-----NMMAFSPYMGTFSPEDVPGLATLVLVEQM 754
Db 666 AVPSQEVRRSAEQLAALWDEQHVSELYAPNLVTEALAEAAEAGRTLPLALRHTAQAGEAL 725
QY 755 SSSVN-----AIWAPKLQLLNGYQSSSSICFASNMSTEP-----NNMGRAY-GAHSW 803
Db 726 TLTRTVREFAAAVPGQLNHHYGPAB-THVMTGTALPEDPAWSEHAPLGRPVSGARVYV 784
QY 804 IDPNDINRLVPI--GAVGELVIESPGIARDYIVPPPEKSPFETD--IPSWYPANTFPDG 859
Db 785 LD-----SALRPVAGVTGELYLAGAGVSRGYNLRPVLTAEFRVADPYAFS-----PG 832
QY 860 AKLYRTGDLARYASDGSIVCLGRIDSOVKIRGORVELGATETHLRQOMDDTLIVVEA-- 917
Db 833 ARMYRTGDLGRWNRGELEFAGRADHQVKIRGPRIEPEIEAAL-----TDLPAVARAAV 887
QY 918 TRQSOSANSTSLAFILGSSYFGNRPDSDAHILDHDKATKAINKLEOVLPHRSPSYICM 977
Db 888 VREDRPGDKRLVAYAV-----PAGEGL---DAA-AVRSRLARTLPDFWMPAAIVAL 935
QY 978 LELPRATGKIDRRIRIMGKOILDQTOGAIVQAPAPIPVFADT--AAKLHSIWQSL 1035
Db 936 DALPLTPNGKLDRAALPA-----PQSTAVQRTSPRPREELITTLFAVLK---LPRV 985
QY 1036 GIDPATVNGATFELGSGNSITAIVKVNARSY-GMDLKVSNIYQHPTLAGISAVYKGD 1094
Db 986 GIDD-----NFFDLGGHSLATRLISRTVLGAETILRDLFEAPTAVAGL-----GER 1033
QY 1095 LSY-----TLIPKSTHEGPEQSYSGRLWFLDQDLVGLSLWLYPVAVMRGPVNVDAL 1148
Db 1034 LQASGARPALPAERPE-RLPLSHAQRRLWFLGRUDGNSTYNIPALRSLGSLDTAAL 1092
QY 1149 RRAALAEQRHETLRTTFEDQGVQ-----IVHEKLSSEMKVIDLCSGL 1195
Db 1093 RAALADLVARHETLRTYPSHDGEYQHIVAPADATPALTVVH---TEHRLA----- 1142
QY 1196 DPEVLNQOTTFNLSAGWRATLRLGEDDHILTYVHHIISDGSIDVLRDLNOL 1255
Db 1143 ---EQLAEATAPYDUTTELPURATFLRGEDEHVLVLLHHITAGDWSLAPLTDLAAA 1199
QY 1256 YSAALKSDPLSALTPLQYSDFAKWOK-----DQFIEQKOLNWKOLKDS 1308
Db 1200 Y-AARDGNAP--GFAPLPQVADYTLWQRLGLGDADDTSLQAQALHWRQAL-BGRPA 1255
QY 1309 --KIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFNCNEHNTTSFVLLAAFAAAHYRLT 1366
Db 1256 HLELPCHDPRPATAVTHRGATVPFHIDAGLHEKLTALAKACDSSLFVLLQAAFAALLTRHG 1315
QY 1367 AVEDAVIGTPIANRNPELEIDIGCFVNTQCMRINDHHDHDTGTLINQVATTAFAENE 1426
Db 1316 AGTDIPVGSVAGRTDDALDVLGVFVNTLVLTDTSGDPTFRELVARVQRQDLAAATHQ 1375
QY 1427 DIPFERVVALQPSRDLSTPLAQILFAVHSQ--KDLGRFEOGLSVVPVPSK-AYTRF 1483
Db 1376 DMPEKLVEBNP-ERSLARNPLFOVVLALQSMAPADLA---LPGLDVAAEPPRVGFAKF 1431
QY 1484 DMFEHLFQETDS-----LKGSVNFADLFEKMTVENVRVFEFELNGLQSSRTPVSTL 1537
Db 1432 DLGLAVIEHTADGTRAGIRGDWEYSTELFERCTVEALQSLRVLRLDADADDPERTIGSV 1491

QY 1538 PLTDGIVTLEKLDVLNVKHVDYPRESSLADYFQTSAYPDSLAVVYDSSCLRTYTELDRQ 1597
Db 1492 DLLDATERHRLVGLNTDGPLPEERTLALFQQAATPDATALYMGORSLSYIYAEALDAR 1551
QY 1598 SDILAGWLRRSRMPAETLVAFAPRSCETIVAFGVLKANLAYLPLDVRSPSARVODILS 1657
Db 1552 ANRLARHLVALGAGPEQIIVALQPLSRSLDLVTAVLAVMKSGAAYLPVDPDPPIERIAHMA 1611
QY 1658 GLSGPTIVLIGHDTAPDTEVTNVEFRIPDALNSNADGFEVIEHDSKTPSATS-LAYV 1716
Db 1612 D-ARPALV--DAIPDDTTLAAADSRLLTDA-----DRSAPLLPAHPAYV 1653
QY 1717 LYTSGTGRPGKGMIEHRIIVTISGCIIP-NYPSETRMAHMAITAFDGASYIYISALLF 1775
Db 1654 IYVTSGTGAPGVVVAHRSLSAATVPAQAAGLGTSHRVLNFASISFDAAWELTSALFT 1713
QY 1776 GRTLVCVDYMTLIDARALKDVFREHVNNAASHVTSQDVP-LRVPRRLSRTLMFFELV 1834
Db 1714 GAGLVADADDLLPGPSLARLVHHRHITLIALPPSALPALPGALPPGTD-----LIVA 1767
QY 1835 TDSTAPDALDAQGLYQVOCYNGYCTENGVMSTIYPIDSTESFINGV-PIGRALNNSGA 1893
Db 1768 GDATAFD--QAARFAPGRWVNAVGLTETVTCATM---SEPATGAGPPIGRPVAHARV 1821
QY 1894 YVVDPEQQLVGIVMGELVVTGDLARGYSK-ALDENRFV-----HITVNDOTVKAYRTG 1948
Db 1822 YVLDERLRPVPGVTGEMTVSGAGVARGYLHRPALTAQRFPVDPYALLFGETGTRMYRTG 1881
QY 1949 DRVRYRIGIGLLEFFGCRMDTQKIRGNRIESABIEAALLRDSVRDAAVVLQNEOQAP 2008
Db 1882 DLARLD-ADGRHLFAGRAQOVKIRGFIEPGIEGTIVLTAHPAAGAVIAREDTFGDKQ 1940
QY 2009 ILGFVVAD-HDISENDKGSANQVEGWQHFESGMSYDICEIDPS-TIGSDFKGTSMYD 2066
Db 1941 LVAYLTRDTTHAAPPQDLGTGOVDRWQTYDS-----LYEAPGRAGFEDFSWNSSYT 1995
QY 2067 GSOIDDEHWEHUGETTRTLHDNRSLG----NVLEIGTSGMILFNLDSKLESYVLEPS 2122
Db 1996 ABEIPLLEEREWRAATV-----DRVLALPRRVLEICGCTGLLSQVAPHTTEYRGTDLS 2050
QY 2123 RSAAAFVNKATESIPSLAGKAKVQVCTATDIGOVDLHP--DLVVLNVIOYFSPSEYL 2179
Db 2051 RSVVARLAHLGRPDADKVTVDARAA---HETDOLPKGRFTIIVLNSVAQIFPDARYL 2107
QY 2180 AETADTLIHL--PNVQRIFFGVRSOATNEHFLAARAIHTLGNATKDDVRQKMAELED 2237
Db 2108 AGILHRAAELLAPG-GTIFLGDILRNLTLHAFRTAVELRRAGFPADPAAVRRAVEQSLAT 2166
QY 2238 EBELLVEPAFFTSKDRFFGLVEHVEILPKNMEAVNELSAYRYAAVHVHVGSLGDELVL 2297
Db 2167 EXELLDDPDDFTDLVDRDPHLT-RADITLRTAAHHNEMSRHRYDVTLLH-RG--GDS---P 2219
QY 2298 VEKDDWIDFOANQLNQLSLGDLKSSDAATMAVSKIPFEITAFERQVAVSLNSNIDEWOL 2357
Db 2220 AFAETVLRGT-----DLDTAGELARLVSGPRPVRTGIPN---ARLTGETRAAQA 2267
QY 2358 STIRSAEGDSSLSVP-----DIFRIAGEAGFRVEVSSARQWSONGALDAVHHCC 2408
Db 2268 LERGSAAACARLEAPAGALDPHDVYAL-----GAVCTWGHG---DDTFDAYA 2314
QY 2409 SOG-----RTLNVFPDHLRGSDLLTNRPLOKRNRIAEVREKRLSLPSYMTPSNI 2463
Db 2315 GDCATYRPTRGQP-----LANDPARAEQDAALVELRDLVAORLPAHPAPAA 2364
QY 2464 VVLDKMLPNAKGVDRKELSRRAKVVYPKQTAAPLPTFFIPESEVILCEATEVFCMKVD 2523
Db 2365 VLLDRPLSANGKLDL-----DALPAP----- 2386
QY 2524 ITDHFENLGGHSLATKLISDRQLKVRITVKDVEDHPVADLASVIRQGLGLOQPVSD 2583
Db 2387 ----- 2386

Db 887 VIVEKDESGQV-----LYAYLV-----ERELTVAELREFLGRILPSYMP 928
Qy 972 SFYICMLPRTATGKIDRRRLRIMGKDILDKOTQGAIVQAPAPIPVFADTAAKLSIW 1031
Db 929 SFFIRLAEIPLTANGKVERKKL-----PRPAGAVVGTAYAAPO-NEIEAKLAEIW 978
Qy 1032 VQSLGIDPATVNGVATFELGGNSITAIAKW-NMARSVGMDLKVSNTIYQHTLAGISAVV 1090
Db 979 QQVILGI--SQVGIDHDPFDLGGHSLKAMTVVFOVSKALEVELPKALFEHPTVAELARFL 1036
Qy 1091 -KGDLPSYTLI-PKSTHE-GPVEQSYSGRLWFLDQDVGSLWLYPIYAVMRGPNVDA 1147
Db 1037 SRSEKTEYTAIQVAAQEFYV--SSAQKRWYILQQFEGNCISYNTSGAILLEKLDYAR 1094
Qy 1148 LRRALAALEQRHETLRTTFEDQDGVQIVHEKLEEMKVIDICGSLDPEFVLNORQTT 1207
Db 1095 FASAVOQLAERHEALRTSPHRIDGEVPQKVHEEVEPLFMLE--APEDQAEKIMREVR 1151
Qy 1208 PFNLSEAGWRATLLRIGEDDHILTIVMHHIISDGSIDVLRDLNOLYSAALKDSKPL 1267
Db 1152 PFDLGVAPLMRTGLLKIGDKRHLFLDDMHIIISDGVSSQILLREFAELYQCA-----1203
Qy 1268 SALTPLPIQXDFAKWQKQDFIQ--EKOLNYYKKQKLDSSP-AKIPTDFARALLSGDA 1324
Db 1204 -DLOPLSLQKDFAAWONELFQYAYKQEQHNLNTFADEIPLNLPTDPRPSVQSFAG 1262
Qy 1325 GCYHVITIDGELYQSLRAFNCNEHTTSPVLLAAAFRAAHYRLTAVEDAVIGTPIANRPE 1384
Db 1263 DLVFAAGKELLERLOQVASETGTYLMLLAAYNVLLSKYTQEDIIIVGTPVAGRSHAD 1322
Qy 1385 LEDIICGFVNTQMRNIDHDTFGILINOVKATTTAAFNEDIPPERVVSALQPSRDL 1444
Db 1323 VENIMGIFVNTALRNQPASSKTFQAFLOEVQNALAAAYDQDYPPFEELVEKLAI-QROI 1381
Qy 1445 SSTPLAQILFAVHSQDKLGRKFQGLSEVPVPSKAY-----TREDMEHFLFOETSLKG 1498
Db 1382 SRNPLEDTLESNANO-----QSLAIELTASPYELFNKISKFDLALNASESPADIQF 1435
Qy 1499 SVNFADLFKMETVENVVRVFFELRLNGLOSSRTPVSILPLTDGIVTLEKLDVL---NVK 1555
Db 1436 QLTFAATKLFKKEVVERMARHYLEILR---WISQPTASLADIDMTEAEKRTLLLVNDT 1492
Qy 1556 HVDYPRESSLADYFQVSAYPDSLAVVDSSCRLTYTDLQSDILAGWLRSSMAETL 1615
Db 1493 FVERTAATLHQLVEQAAARTPEVAVVVEEYALTYRELNARANQALRLRSHGTGPTDL 1552
Qy 1616 VAYFAPRSCTEIVAFFGVLKANLAYPLDVRSPSARVQDILSLSGTIVLIGH--DTAP 1673
Db 1553 IGIWDRSPGMVGVMLAVLKAGAYTIDPSYPERIQYMLSDSQAPILLTQRHLQELAA 1612
Qy 1674 PDEVTNVEFVRTDALNSNADGFEVIEHDSKPSATSLAYVLYTSGSTGRPKGYMIEH 1733
Db 1613 YQGEIIVDDEAYTGA-DTNLDNV-----AGKDDLAYVYVYSGSTGNPKGVNISH 1662
Qy 1734 RVLIIRVTSCCIPNYP--SETRMAMHATIAFDGASYEIYSALLFGRTLVC-----VD 1783
Db 1663 QALCNHML-WMRTFFPLTTEDAVLQKTPFSFASVWVEFYLPITGGOLVLAKPDGHRDIA 1721
Qy 1784 YMTTLARALKQVFFREHVNAAASHVTSQSDQVPLRVPRRLSRFLMFFFLVVTDTSTAPDAL 1843
Db 1722 YMTRL-----IRDEKITTLQWYPSLLDLVMTDPGWSACTSLQRFVCGGALTALPV 1772
Qy 1844 DAQGLYQGVQCYNGYPTENGVMSTIYPIDSTESFNGVPIGRALNSGAYVVDPEQQLV 1903
Db 1773 SRFYETQQAQLINLYGTETTTIDATYWPQROQY--SAIPIGKPIDNVRVYVNASNQLQ 1831
Qy 1904 GIGVMGELVYTGDLARGY-SKALDENREHVHTVNDQIVK----AYRTGDRVRYRIGDG 1958
Db 1832 PUVAGELIAGDLARGYWQREELTKASFV-----DNPFEPGNTMYRTGDMVRY-LPDG 1885
Qy 1959 LIEFFGRMTQKIRGNRIESABIEAALLRDSRDAAVVLQONEDQAEIILGFVVADHD 2018
Db 1886 HIEYLGRIDHVKIRGHRIELGEIATLLQHEAVKAVVMARQDGKGQNSLAYVVAEQD 1945
Qy 2019 -----HSENDKGOSA-----N 2029
Db 1946 IQTAELRTYLSATLPAYWPSAFVLEQLPLSANGKYDRKALPOPEDAASAAVYVAPRN 2005
Qy 2030 QVE-----GQWDHF-----2038
Db 2006 EWEAKLAAIWESVLGVEPIGVHDHFFELGGHSLKAMHVISILLQORSFOVDVPLKVFLESPT 2065
Qy 2039 -----ESCMYSDI-----G 2047
Db 2066 IAGLAPLVAARKGTYTAIPVEKQYYPVSAQKRMFILOQMBGAGISYNMPCFMYLDG 2125
Qy 2048 EIDPSTIGTSFKGWTSMYDGSQIDFDEMHEWLAGEETTRTLHDNRSLG-----2093
Db 2126 KLQTERLOQALKSLVQRHESLRTSF---HSVQGETVQRVHDDVDLAIISFGEATEAETROI 2182
Qy 2094 ---NVLEITGT---SGMI-----LNLQSR-----LESYVGL-----2119
Db 2183 AEOFIQPDFLGTAPLLRAGLIKLAPERHLFMDLHHIWDGVSIGILLIEFAQLYHGEEL 2242
Qy 2120 -----EPSRSAAAFVNKATESIP-----SLACKAK 2144
Db 2243 PALRIQYKDFAKWQDMFQTEFAEQAYWNLFTGEIPVNLNPTDYPSPVSKSFAGDRF 2302
Qy 2145 VOVGATADI-GQVDDDLHPD-----LVVLNSVIOYFSPSEYLAETADTLIHLNVQR 2194
Db 2303 V-FSGGTALPKOLHQLAQETGTTLYWVLLAAYNVLLSKYSRQE-----DIIVGAPTAGR 2355
Qy 2195 -----IFRGD---VRSQ-----ATNEHFLAARAIHTLGNKATKDDVRQKMAELBDE 2238
Db 2356 SHAETESIVGMFVNTLALRNEPAGGKTFRDLFAEVKINTLGAFEHQDYPLDELVDKLDQ 2415
Qy 2239 EELLVPEAPFTSL---KDRFPGVLEHVEITLP-----2266
Db 2416 ROLSRNPLFTVFIQNMEOKPFEMEBOLITTPYSAEVKQAKFDLSLEAYEENABIIFSLD 2475
Qy 2267 -----2266
Db 2476 YSTKLSFRETIERKIATHFIQILRAVIAEPEMLSEITMLTAEKQRLLVDFNGAHKDFPQ 2535
Qy 2267 -KNWEAVNELSAYR--YAAVVHVRG-----SIGDELVL-----VEKDWDIDPOAN 2309
Db 2536 NKTLOALFEQAEKSPQATAVEISGQPSLSYQELNERANQLAATLRERGVDQDPQVGM 2595
Qy 2310 QLNOKSIGDL--LKSSDAAIMAVSKIPFEITAF-----PROVVASLNSNT---2352
Db 2596 RSVEMVVGILAILKAGAYVIDPEYPERVAVYMLTDCQARLVLTOKHLGAKLGSSVTAE 2655
Qy 2353 -----DEWOLSTIRSAEGDSSLS-VPDIFRIAGEAG-----F 2384
Db 2656 CLYLDDESNGVHRSNLQPIINTASDLAYIITSGTGKPKGVMMVHRGIVNNVLWKAAY 2715
Qy 2385 RVEVSSARQWSQALDAV---FHHCCSQGRTLVNFTDHLRG-----SDLL 2429
Db 2716 QMKVGRSLLSUSFAFADFVLSFTFVLSGATVV-LAEDEEAKDPVSLKLLIAASRCTLM 2774
Qy 2430 TN-----RPLQ--RLQNRRIAEVRELRSLPSVMI-----2459
Db 2775 TGVPSLFQALLECSTPADIRPLQTVILGGEKITAQLVEKCKQLNPLDIVINEYGPTESSV 2834
Qy 2460 -----PSNIVLQK-----MPLNANGK--VDRKELSR---RAKV 2488
Db 2835 VATWORLAGPDAAITIGRPANTSLVIVNOYHOLQPIGVVGEICIGRGLARGYWNKPAL 2894
Qy 2489 VPKQQTAAFLPT-----FPISEVEVI-LCEATEVEFGMKVDITD-----2526
Db 2895 TEKEFVSHPFAGERMYKTGDLGKWLPLDGTIEYIGRIDEQVKVGRYRIEIGETIESALLAA 2954
Qy 2527 -----HFNLCGHSLA-----TKLSRDIQRLKVRITVKTVDHFPVAD 2566
Db 2955 EKILTAVVVYEDLQGSALAAAYFTADEQLDVTKLWSHLSKRLPSTY-----IPAHFQOLD 3010

Db 2476 HTTENOMEYKDTISRLNHLKRLERQIIKHCTOQTESERTPSDYGDTNISLAB----- 2529

Qy 2700 ASEIKPOLEQEIQDIYPTQMOKAFLE----DHTT-----ARPRPVPFYIDFPSTSEP 2751

Db 2530 LEEIKG--KYRSALEKIYPLANMKGMLFAIEDHTSDAYFOQIVMDIEGYVD-PAILEA 2586

Qy 2752 DAAGLIACSLNHLDIRFVFAEASGELYQVVLSCDLDPQIVETEDNI-----NTAT 2806

Db 2587 SFNDIMKRHEIL-----RASYE-YEIV-----EPRQIIENRSIDETIFYENIAK 2629

Qy 2807 NE-----FLDEFAPKPEVRLGHPLIRFTIILK-----QTKSMRVIMRISHALVDGLSLEH 2854

Db 2630 SRAOQOEMFIERLLNEDRKKGFGLSKVDLWRAVLLKTAERSYRLVWSHHILLDGCWCLGI 2689

Qy 2855 VVRKLHMLY----NGRS--LLPHQFQSRMYQY-TADGRESHGFWRDVI-----ONTPTMTI 2903

Db 2690 IMRELFIYENRMNKGASPLKETPKYSDYIKWLERQDOBEARQYWRVYLGKYEQAQLPT 2749

Qy 2904 LSDTVVDGNDATCKALHLSKIYNIPSOVLGSSNIITQATVFNAAACALVLSRESKDV 2963

Db 2750 LTRKKSSRYDRREKVIHLSKQLTKQKEL-AAKNSVTLTHTVIQTWGLMLTRYTKIDDV 2808

Qy 2964 VFRIGVSGROGLPVEYODIYGPCTNAVPPVRAHI-ESSDYNOLLHDIDQVLLSLPH---- 3018

Db 2809 VFTGVVSGRANGVDGIEDMIGLFINTIPTIRNEQARFNDCLQKQVEDAIQSNRYNYM 2868

Qy 3019 --ETIGSDUKRNCRTDMPAITNFSCITVYHFEYHPESQFQOORVMGV----- 3066

Db 2869 LAEQVALLSLKKDLIDH-----ILVPENYE-ADEQDFEESQMKTFKFNVEISAAEQ 2918

Qy 3067 LTKF-----VNIEMDEPLVDLAGEVEPDGAGLKVTYAKTOLFGRK--RVE-3112

Db 2919 SITAFMSVTPGELTLVLTYDNGVYDRDIINNIE---GHIKRVAEQVTANENRKTAIED 2975

Qy 3113 HLLIEVSKT 3121

Db 2976 MIAEERKT 2984

RESULT 7

JX0340

N;Alternate names: gene Grs2 protein

C;Species: Bacillus brevis

C;Date: 22-Apr-1995 #sequence_revision 26-May-1995 #text_change 03-Nov-2000

C;Accession: JX0340

R;Saito, F.; Horii, K.; Kanda, M.; Kurotsu, T.; Saito, Y.

J. Biochem. 116, 357-367, 1994

A;Title: Entire nucleotide sequence for Bacillus brevis Nagano Grs2 gene encoding gramicidin S synthase 2 - Bacillus brevis

A;Reference number: JX0340; MUID:95122465; PMID:7822255

A;Accession: JX0340

A;Molecule type: DNA

A;Residues: 1-4450 <SAI>

A;Cross-references: DDBJ:D29676; NID:g473695; PIDN:BAA06146.1; PID:g511490

C;Comment: This enzyme catalyzes activation and combination of four constituent amino acids to form gramicidin S. The enzyme contains 4 domains with a mean of 1042 amino acid residues containing 137 gaps.

C;Superfamily: alpha-aminoacyl-tRNA synthetase; acetyl-CoA ligase homolog

C;Keywords: carrier protein; phosphopantetheine; phosphoprotein

F;514-956/Domain: acetate-CoA ligase homolog <ACLI>

F;974-1042/Domain: acyl carrier protein homolog <ACPL>

F;1552-1991/Domain: acetate-CoA ligase homolog <ACL2>

F;2009-2077/Domain: acyl carrier protein homolog <ACP2>

F;2588-3036/Domain: acetate-CoA ligase homolog <ACL3>

F;3054-3122/Domain: acyl carrier protein homolog <ACP3>

F;3634-4074/Domain: acetate-CoA ligase homolog <ACL4>

F;4092-4160/Domain: acyl carrier protein homolog <ACP4>

F;1006,2041,3086,4124/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 12.0%; Score 1933; DB 2; Length 4450;

Best Local Similarity 22.0%; Pred. No. 1.le-103;

Matches 845; Conservative 587; Mismatches 1306; Indels 1098; Gaps 137;

Qy 67 KQSAIGHAVYDVPTDIDISRFALAWKEIVN---QTPALRAFAPTSDSGKTS----- 114

Db 19 QEGMLFHALDKOKNAHLVQMSIAIEGIVDELLSESLNILLIDRYDVFRTTELHEKIKQP 78

Qy 115 -QVILKDSFVSWCWSSSSPDEVRDEAAAA---ASGPRCNRFLVED-----MQT 163

Db 79 LOWVLKERPY--QLQFKDISLDEEKREQAIEQKYQDGETV--FDLTRDLRMAVAFQT 134

Qy 164 KKC--QLVHTFSLHVDVTFQQRVLSRVFAAYKHEKDTHRPETPESDADTDTSQSISVV 221

Db 135 GKVNQMINSEFHLLMDGCFNFIIFNDLFNIIYLSLKE---KKPLQLEAVQPKQFKWL 190

Qy 222 SMCEDNAVSAATHFWQTHLDNLNASVPHLSHDLMVNPNTT--TAEHRTITFP----- 271

Db 191 E---KDKQKQALRYKWEHLNMYDQSV--TLPKKAAAINNTTYEPAQFPAFADKVLTOQLL 245

Qy 272 --LSQKALSNAICRTALSTILLSRYTHSDEALFGAVTEQSLPDKHYLAD--GTIQTIVAP 327

Db 246 RIANQSOVTLNIVFQTIWIGIVLOKYNSTNDVYGVSV--SGRPSSEISGIEKMGVGLFINTLP 304

Qy 328 LRVHCOSNLRASDVMDAISS---YDRLGHLAPGLRDINRTGDNCSAACDFQVL--- 380

Db 305 LRITQKQDSFIELVKTIVHQNVLFSQOHEY---FPLYEQNHTELKONLIDHINVIENYP 361

Qy 381 LVTDGSHVNNINGFLQO-----ITESHFMPCNNRALLLHCQMESSGALL---VAYYDHN 433

Db 362 LVBELOK-----NSIMQKVGFTVRDVKMFETNYDMTVMLPRDEISVLDYNAAYVDID 416

Qy 434 VIDSLOTTRLLQFGHLIK---CL-QSPDLSSMAEVLNLTAYEDRAE--IESWNSQPLEVQ 488

Db 417 FIKRIE-----GHKKEVALCVANNPHVL--VQDVPLLTQKQHLVVELHDSITEYP 466

Qy 489 DTLIHHEMLKAVSHSPTKTAIQAWDGDWTVSELDNVSSRLAVHVKSLGRRAQAAIIPYVF 548

Db 467 DKTIIHQLETFQVEKTPHVAVVFEDKVTYRELHERSNQLARLREKGVK--KESIIIGMM 525

Qy 549 ESKWVIASMLAVLKSGNAFTLIDPNDPPARTAQVVTQTRATVALTSKLHRETQVKLVGR 608

Db 526 ERSVEMIVGLTILGILKAGGAFVPIPEYKPERIGYMLDSVR--LVLTQR--HLKOKFAFKE 582

Qy 609 CVVDDDELQSVASDDFSLTKSQDLAVYVFTSGTGDGPKGIMIEHRAFSKALRFGAS 668

Db 583 TIVIEDPSI-SHELTEIDYINSEDLFYIYISGTTGKPKGVMLHKNIVNLLHFTFEK 641

Qy 669 LGINSTRALQFTHAFGACLLIMTTLNGG--CVCIPSDDDR--MNSIPSFIRNYNVMW 726

Db 642 TNINFSDKVLQYTTCSFDCYQEIFSTLLSGGQLYLRKETQRDVEQLFDLVKRENIE-V 700

Qy 727 MATP-----SYMGTFSPEVDYVGLATLVLGEOM---SSSVNAIWAQKLOLLNGY 772

Db 701 LSPFVAFLPKIFNREFINRF-PTCVKHIIT---AGEQLVNVNEFKRYLHEHNVHLNHY 756

Qy 773 GQSESSI-CFASNMTSEPNM---GRAVGAHSHWIDPNDINRLVPICAVGELVIESPGI 828

Db 757 GPSETHVVTYITINPAEIPPELPICKPI-SNTWIYILDOEQQLQPGQIVGELYISCANV 815

Qy 829 ARDYIVPPPEKSPFTTDPISWYTPANTFPDGAKYRTGDLARYASDCSIVCLGRIDSQVK 888

Db 816 GRGYLNNQELTAEKFFAD-----PFRPNERMYRTGDLARWLPDGNIEFLGRADHGVK 867

Qy 889 IRGQVELGAIETHLRQOMPDDLTIVVEATKRSQSANSTSLIAFLIGSSVFGNRPSDAHI 948

Db 868 IRGRIELGEIEAQLLNCCKGVKEAVVID---KADDKGGKYLCAIVV-----MEVEV 915

Qy 949 LDHDATAKAINIKLEQVLRHSIPSFYICMLELPRTATGKIDRRRLRMGDKILDQKTOGA 1008

Db 916 NDSE---LREYLGKALPDYMIPEFVPLDQLPLTPNGKIDRSL-----PNLEGI 962

Qy 1009 IVQOAPAPIVFADTAALKHSHIWSQLGIDPATVNVGATFFELGGNSITAIKMNW-MARS 1067

Db 963 VNTNAKYVVVPT-NELEEKLAKEIWEVLGI--SQIGIQDNFFSLGSHSLKAITLISRMKE 1019

Qy 1068 VGMDLKVSNIYQHTPLAGISAVVKG--DPLSYTLIPKSTHSGPV-EQSY-----SQGLRWF 1120

Db 1020 CNDIPLRLLFEAPTIOBISNYINGAKKESVAIQ-----PVPEQEYYPVSSVQKRMFI 1073
QY 1121 LDOLDVGLWLYIPVAVMRGPVNDALRRALALEQRHETLRTTFEDODGVQIVHEK 1180
Db 1074 LNEFDRSTAYNLPGVWFLDGLKLVNRQLEAAVKLVERHEALRTFSHSINGEPVORVHON 1133
QY 1181 LSEEMKYIDLCSOLDPEVLNQEQTPFNLSSEAGWRATLLRLGEDDHILTIYVMHHIS 1240
Db 1134 VELQIAYSE---STEDQVERIIAEBFQMPFALEVPALLRVLGLVLEAERHLFIMDMHHIS 1190
QY 1241 DGWSDIVLRDLNQLYSAAKDSKPLSALPLPIQISDFAKW-----OKDOFTIEQKOL 1295
Db 1191 DGVSQIMIQIADLYKE-----KELEPTLGIQYKDFVTVMHNRLLQSDVI---EKQE 1238
QY 1296 NYWKKQLKDDSPA-KIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFECNEHNTTSFVVL 1354
Db 1239 AYWLNVFTEETPVNLPTDYPRPTIOSFGDGRFTSTGKQLMDDLYKVATETGTLMYVL 1298
QY 1355 LAAPRAAHYRLTAVEDAVIGTPIANRRPELEDIIGCFVNTQCMRINIDHDHDTGTLNQ 1414
Db 1299 LAAYNVLSKYSQODDIWVGTPAGRSHADVENMLGMPVNTLAIRSLNNEDTFKDFLAN 1358
QY 1415 VKATTAAAFENEDIPFERVWSALOPG-SROLSSTPLAOLIFAVHSQKDLGRFKQGLESV 1473
Db 1359 VKOTALHAYENPDYFDTLVEKL--GIQRDLRSRNPFTDMFVLQN-TDRKSFEVEQITIT 1415
QY 1474 P-VPSKATREDMEFHLPQEDSLKGSVNADELFKMETVENVVRFEILLRNGLOSSRT 1532
Db 1416 PYVNSRHSKDFLEVESEONEILLCEYCTKLTFTKTVERMAGHFIQIILHAIVGN--- 1472
QY 1533 PVSILPTDGTVITLEKLDVL---NVKHVDYPRESSLADVFQTVQSAYPDLSLAVDSSCRL 1589
Db 1473 PTIIISEILSEEEKHILFEENDTKTTPHMQTIOGLFEEQVEKTPDHVAVGWKQDAL 1532
QY 1590 TYTELDRSDILAGLWRRSPMAETLVAVAPRSCETTIVAFGVILKANLAYLPDVRSPS 1649
Db 1533 TYRELNERANQVARYLROKGVQOPNIVGLLVERSPEMLVGIMGLKAGGAYLPDPEYPA 1592
QY 1650 ARVODILSGLSGPTVLIGHDTAPDIEVTNVEF---VRIDALNDSNADGFEVIEHST 1706
Db 1593 DRISYMIQD-CGVRIML---TQOHLLSLWDEDFCVILDEDSLYKGDSSNLAPVN--- 1643
QY 1707 KPSATSLAYLYTSGSTGRKPGVMIEHRIIIVTSTGCPINYPSETRMAHMATIAFDGAS 1766
Db 1644 --QAGDLAYIMVTSGSTCKPGVMVHRNVRILVKNTNVQVREDDRIIQTGAIGFDALT 1701
QY 1767 YEIYSALLFGRTLVCVDYMTILDARALKDVFREHVNAAASHVTSSQDVPLRVPRRLSRT 1826
Db 1702 FEVFGSLHGAELPVTKDVLDDAEKHLKFL-----QANQITIMWLTSPLF--NQLSQG 1753
QY 1827 LMFFF-----LVVTDSTAPDAL-DAOGLYGVOCYNGYPTENGVMSTIYPIDSTESI 1879
Db 1754 TEEMFAGRLSLIVGSDALSPKHNNVKKCPNLWNGYGTENTTFTSCFLID--KEYD 1811
QY 1880 NGVPTGRALNNSGAYVVDPEOOLVGIVGMELVWTGDLGARYSDK-ALDENRFVHITVN 1938
Db 1812 DNIPGKAISNTVYIMDRYGOLOPVGVGELCVGGDGVARGYMNQNPATEEKFVNPFA 1871
QY 1939 DQTKAYRTGBRVYRIGDGLIEFGRMDTFKIRGNRIESAETAEALLRSSVYDAV 1998
Db 1872 PGE-RMYRTGDLAR-LPDGTIEYLGRIDQOQVKIRYRIEPEIETLLVKKKKYESVIM 1929
QY 1999 LQONEDOAPETILGFVVADHD-----HSENDK--- 2024
Db 1930 VVEDNNGOKALCAYVPEEVTSELREYIAKELPVYVMPAFVQIEOMPLTONGKVNRS 1989
QY 2025 -----GOSANQVE-----GWQDHF----- 2038
Db 1990 ALPKPDGEGGTATEVAPSSDIEMKLAETIHNVLGVNKGIVLNDFFELGGLHSLRAMTIS 2049
QY 2039 -----ESGMYSIDGELDPST-----IG 2055
Db 2055 -----PEPDGSIISIGTIEYVAPRTMLEKLEIWKDVLGLQVRGIIHDDFTTIGGHSUKA 3089
Db 2050 QVHKFEFVELPLKVLFETPTTISALAQYIADGKGMYLAIQVPTPODYYPVSSAQKRMVIL 2109
QY 2056 SDFKWTSMYDGSQIDFDE-----MHEWL-----GETRTRLHONR 2090
Db 2110 YEFEGAGITYNPNVWFIETGKLDYORFEYATKSLINRHEALRTSPYSLNGEPVQVHONV 2169
QY 2091 SL-----GNVLEIG-----TSGSMILFNLDSR----- 2112
Db 2170 ELQIAYSEAKEDETBQIVESFQPPDLEIAPALRVGLKASDRHLFMDMHHSIDGVS 2229
QY 2113 -----LESYVGLGPS-----RSAAA-----FVNKATESIPSL-- 2139
Db 2230 MCIITKEIADLYKKGELAEHLHIQYKDFAVQWQNEFQSAALEKQKTYWLTFAEDIPVLNL 2289
QY 2140 -----AGRAK 2144
Db 2290 STDYPRPTIOSFEGDIVTFSAGKQALAEELKRLATETGTYLMLLLAAYNVLHLKXSGQEE 2349
QY 2145 VOVGTA-----TDIGQVDDLLPD-----LVLVNSVIOYFPSEYLA 2180
Db 2350 IIVGTPAGRSHADVENIVGMFVNTLALKNTPIAVRTFHEFLLEKQNALEAFENQDYPF 2409
QY 2181 E-----IADTLHLPNV-QRIFFG-----DVR 2201
Db 2410 ENLIEKLOVRDLRSRNPFDTFMSLSNIDEQVEIGIEGLSFSPYEMOYIAKFDISFDIL 2469
QY 2202 SOATNEHF-----LAARAIHTLGNATKDDVROKMAEL-----EDMEBEL 2241
Db 2470 EKODDIOFYFNCTNLFKKEITERTLATHFMHILOEIVINPEI--KLCEINMLSEEQORV 2527
QY 2242 LVE-----PAFFTSLKDRFPGVHEVHEILPKMNAEYVNELSAYRY----- 2280
Db 2528 LYDFNGTDATVATNKFHELFEQVEKTPDHIAVIDEREKLSYQELNAKANLARVLROK 2587
QY 2281 -----AAVHVVRG 2288
Db 2588 GVQPNMVGIMVDRSLDMIVGLMVLKAGGAYVPIDIDYPOERISYMMEDSGAALLTQO 2647
QY 2289 SLGDBEL-----VLPVEKDDWDIFOANQLN----- 2312
Db 2648 KLTQOIAFSGDILYLDQEEWLEHEASLEPIARPQDIAYIIVTSGTTPKPGVMIEHOSY 2707
QY 2313 -----OKSLGDL-----LKSSDAIM 2328
Db 2708 VNVAMAKDAYRLDTFPVRLQLMASFADFVSAGDFARALLTGQQLIVCPNEVKMPASLY 2767
QY 2329 AVSKIPFEITAFER--QVVASLNSNIDEMQL--STIRSAEGDSSLSVPDIFRIAGEAGF 2384
Db 2768 AIIK-KYDITITEATPALVIPLIMEYIEOKLDISOLQILIVGSDSCSMEDFKTLVSRFGS 2826
QY 2385 RVEV-----SSARQWSQNGALD-----AVFH 2405
Db 2827 TIRIVNSYGVTEACIDSSYEQPLSSLHVTGTVPIGKPYANKMYIMNQYLOIQPVGVIG 2886
QY 2406 HCCSQR-----TLVNFTPDHHLRSGDLLTNRPLOK----- 2436
Db 2887 ELICIGGAGVARGYLNRPDLTAEKFPNFPVPEKLYRTGDLARWMPDGNVFLGRNDHQV 2946
QY 2437 -LQNRRIAL-----EVRERLRS 2452
Db 2947 KIRGIRIEGETEAOALRKHDSIKEATVIAREDHMKKEYLCAYMVTEGEVNVAEILRAYLAT 3006
QY 2453 LPSYMNIPSNIVLDMPLNANGKVDKRLSRAKRVVPKQQAAPLPTFPISEVEVILCE 2512
Db 3007 DLPAAMIYSYVSLEAMPLTANGKIDKRSK----- 3036
QY 2513 EATEVFGMKVDITDHFHNLGGHSLLATKLISIDORLKVRTVKDVFDPHFADLASVIR 2572
Db 3037 ----- 3036
QY 2573 QGLGLQVPVSDQOQDRSNHMAPRTETEAILCDEFKVLGFQ-VGITDNFFDLGGHSLMA 2631
Db 3037 -----PEPDGSIISIGTIEYVAPRTMLEKLEIWKDVLGLQVRGIIHDDFTTIGGHSUKA 3089


```
Db 1004 FETPTAGLSALDTSCTLRALTARPRPERIPLSYAQORLWFLHOLEGPTATNIPPTL 1063
QY 1138 RMRCFVNVDALRALAALOEORHETLRTTFEDQGVGOIV--HEKLSSEMKVIDLGGSDL 1195
Db 1064 RLCTGLTDALQSFANDLLARHESLRTYTTEDDGAQOIVLPVAVATPPFAVDVAEDV 1123
QY 1196 DPFEVLNQBQOTTPFNLSSEAGWRATLLRLGEDDHLITVMHHIISDGWSIDVLRDLNQL 1255
Db 1124 --AERVAEAAAHAFDGAIEPVRARLFRVSRHVVHLLLVHHIASDAWSRGLAQDLTAA 1181
QY 1256 YSAALKSDKPLSALTPLPIQYSDFAKWQDQFEQE-----KOLNYWKKQLK--SSP 1307
Db 1182 YTARCADD--APAWQPLPVQYADYALWQOILGDDTDPDTLAGRQLAYWKKOQLAGLPEQ 1238
QY 1308 AKIPTFARPALLSGDAGCVHVTIDGELYQSLRAFPCNEHNTTSFVLLAFAFRAAHYRLTA 1367
Db 1239 LDLPTDRPRATADHTGDRVTFTPADLHTLTELARETNTAFWVIOQAALALLTRHGA 1298
QY 1368 VEDAVICTPIANRNRPELEDIGCFVNTQCMRINIDHHDTEGTLINQVKATTTAAFENED 1427
Db 1299 GEDIPICTPVAGRTDDATDHLIGFFVNTLVLRDTSGNPTFRDLLTRVRDOLDLTAYTHQD 1358
QY 1428 IPFERVYSALQPGSRDLSSSTPLAQILFAVHQSQDLGRKFGLESVPVPSKAYT--RFDME 1486
Db 1359 LPFERLVEALNP--TRSLTHHPLFQTMLTLLHNTQGTKTDRFAGLAAEVVASEVSAREFDLS 1417
QY 1487 FHLQOETDS-----LKGSVNFADELKMEVENVVRVFFELRLNGLQSSRTPVSILPLT 1540
Db 1418 FALAHEFGADHSCAGMSGVYIYSTALFDRATVRDLADRLVLLAAVAHPGRSGVQLEIM 1477
QY 1541 DGIWTEKLDVNLKHHVDYPRESSLADVFQTSAYPDSLAVPSSCLRTYTELDQRSDI 1600
Db 1478 DAAERRLVLOQWNTAAE--PPAASVTGLFERQARRSPGATALEFGEVRSVLAELNARANR 1536
QY 1601 LAGLRLRSMAPETLVAVFAPRSCETIVAFPGVLKANLAYLPLDVRSPSARVODTLISLS 1660
Db 1537 LARHLVGAAGAPGRFVALPRSAELVVTLIAVLKSGAAYLPIDPHYPADRVEYMLAD--A 1595
QY 1661 GPTVLIGHDTAPPDIEVTNVEFVRIRDALNDSNADGPEVIEHDPSTKPSATSLVLYTS 1720
Db 1596 GPALTV---TEP---VAEAGLSYGDA--DLGDELGRPVHGHP-----ATYIYS 1639
QY 1721 GSTGRPKGVMTEHVRVITVTSGCCIPNY-----PSETRMAHMATIAFDGASYEYI 1770
Db 1640 GSTGRPKGV-----VPRGALDNFLADMGRFRFTPGSGDRLLAVTTVGFDIAGLEIF 1690
QY 1771 SALLFGRTLVCVDYMTTLDARLKDVFREHVNAASHVT-----SSSQDV 1815
Db 1691 LPLHGAVLVLADEETARDPHALL-----HRVSASGITMVOATPSLMQGVAAVAGDELA 1744
QY 1816 PLRV-----PRRLSRPLMEFFLVVTDSTAPDALDAQGLYQGVQYNGYGTENGVMST 1868
Db 1745 GVRVLGGEALPSELARAL-----TD-----RARSVTNLYGTEATIAT 1784
QY 1869 IYPIDSTESFNGVPIGRALNNSGAYVDPQQLGVIGVMGELVVTGDLARGYSK--AL 1927
Db 1785 --AADVAES--GPVIGRPLANTSAVILDSALRPVPGVPGELYLAGELQAGYHLRALP 1839
QY 1928 DENRFVHITVNDQIVKAYRTGDRVYRIGDGLIEFFGRMDQFKIRGNRIESAELALL 1987
Db 1840 TSERFADYPGAGTRMTRTGDLV--CRRRDGALRYLSRVDOQVKLGRFRIELGEAEALS 1898
QY 1988 RDSSVRDAAVVLQONDOAPEILGFVAVADHDSNKGQSANQVQEGWQDFHESGYSIDG 2047
Db 1899 RHPAVAESAV----- 1908
QY 2048 EIDPSTIGSDPKGWTSMWDGSDQIDFDEHHEWLGETTRTLHDNRSLGNVLEYIGTGSGMILF 2107
Db 1909 -----TVREDRPG----- 1916
QY 2108 NLDGRLESYVGLERPSRAAFVNKATESIPSLAGKAKVQVGTATDIGOVDLHPLDVLVLN 2167
| | | |
```

```
1917 --DRRLGVV----- 1924
2168 SVIOQYFSPSEVLAETADTLIHLPNVORIFFGDRSOATNEHFLAARAHTLTKNATKDDV 2227
1925 ----- 1924
2228 ROKMAELEDMEELLEVPAFFTSLKDRFPGLVHEHVEI LPKNMEAVNELSAYRYAAVVHVR 2287
1925 -----VPKGPE----- 1930
2288 GSLGDELVLVEKDDWIDFQANOLNOKSLGDLCLKSSDRAIMAVSKIIPFEITAFERQVVAS 2347
1931 ----- 1930
2348 LNSNIDEMQLSTIRSSAEGDSSLSVPDIFRIAGEAGRFEVSSARQWSONGALDAVFHHC 2407
1931 -----GPAGTGPVPSAR----- 1943
2408 CSQGTLYNFPDHHHLRGSDLLTNRLORLQNRRIAEVRELRRLSLLPSYMIPIINVLVD 2467
1944 -----ELRERLGRRLRDEFMVPTAFVVLVD 1966
2468 KMLNANGVDORKELSRRAKVVVKQQTAAPLPTFPPISEVEVILCEATEVFGMKVDITDH 2527
1967 ALPLTPNKKLDRK-----ALPAPEY----- 1986
2528 FENLGGHSLLATKLISRIDQRLKVRITVKDVPDHPFADLASVIRQGLQOPVSDGQO 2587
1987 -----DGEVPV 1991
2588 DRSAHMAPRTEATEILCDEFAKVLGF--QVGITDNFDFLGGHSLMATKLAVRIGHRLDVT 2646
1992 GR---GPRDPREELICALFAEVLGVARVGDGDFLGGHSLLATRLVSRIRLTALGVBL 2047
2647 SVKQVEDHPVLQALALDNLVQSKTNEIVGGRMAEYSPFOLLTFDEPEEMASIKPQ 2706
2048 SVRQFEFTPTAGLSGALD-----RAAGARAP-----LAARPRE 2082
2707 LELOEIIQDIYPSTOMQKAFDHTTARPRFPVFPYIDFPSTSEPDAAGLIKAC--ESLVN 2765
2083 R-----VPLSPAQQRLWFLHQFEGPSATYNNPTALRLSGPVDRAALERAIGDVLVR 2133
2766 HLDIFRTVFAESGELYQVWLSC-----LD--LPQVITEDNINTATNEFLIDEFAKEPV 2818
2134 H--ESLRTVFAADDGGSQWVVLPAVRAVGRDLVDVDTAGEVABRVGEAARHAFDLTADIFP 2192
2819 RLGHPLIRFTTIKQTKSMRVIMRISHALYDGLSLEHVVRKLHMLYNGR--SLLP-----P 2871
2193 -----LARLFRVSDTEHV--LLULIHHIAGDGSMAPLARDLTAAYAAACAGAAPDWEPLP 2246
2872 HQFSRYMOYTAD--GRESGH-----GFWRDVIQNTPMTTILSDDTVVVDGNDATCKALH 2921
2247 VOYADYALWQREVLGDESDDPSVAAARQALAYWKDALAGLPEQL---ELPTDRPRPATAGYT 2303
2922 LSKVINIISOVLGRSSNIITQAT-----VFNAACALVLSRESDSKDVVFGVIGSRQGL 2975
2304 GDRIATVPTALHARLTELARSTHSLFVQAAAFATLLTRLGAGEDIPVGTAVAGRND 2363
2976 PVEYODIVGPCNNAVPRVAHIESS--DYNOLLHDIDDOYLLSLPHETIGFSDL-----KR 3028
2364 ATE--DLVGFFVNTVLRTDTCGNPTFFRELLGRVRDELUAAVHQDVPFERLVEALNPAR 2421
3029 NCTDWPEALTNFSCCITYHNFEYHPESQEQORVEMGVLTKEVNTMEMDEP----- 3078
2422 SLAHP-----LYQVMITENN-----TAGAGDRTAPASPDTPDVSGQAGALL 2463
3079 -----LYDLATA--GE--VEPDG--AGLKVTIVIAKTQIFGRKRVRHLEEV 3118
2464 NATRMTAGTGVAKFDLALTFRGHRDPGSGPAGMRGSLEYRTFLDFRETAEVSIRL 2519
```

peptide synthetase [imported] - Nostoc sp. (strain PCC 7120).

C;Species: Nostoc sp.

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C;Accession: A12136

R;Nakazaki, T.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Iriquchi

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An

A;Reference number: AB1807; MUID:2r595285; PMID:11759840

A;Accession: A12136

A;Molecule type: DNA

A;Status: preliminary

A;Residues: 1-2588 <KUR>

A;Cross-references: GB:BA000019; PID:BAR74347.1; PID:gl17131741; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all2648

C;Superfamily: peptide synthetase ppsD; acetate-CoA ligase homology; acyl carrier prote

C;Keywords: carrier protein; phosphopantetheine; phosphoprotein

F;1012/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

F;2059/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 11.8%; Score 1896.5; DB 2; Length 2588;

Best Local Similarity 23.0%; Pred. No. 5.5e-102;

Matches 736; Conservative 480; Mismatches 1051; Indels 937; Gaps 100;

41 LQSSRIEAKPCPTFPOLDMDICNALDKQSAIGHAVYD-----VPTDIDISRF 87

1 MESSIID-IYELSPMQAGM-----LFHTLYAPNSGVYFEQRCSLIRGNLNLPAF 48

88 ALAWKETVNTGTPALRAFAFTSDSGTKSOVILKDSVFESW--MCWSSSSSDPEVVRDEAAA 145

49 RQAWQVWEHSHVLTAFHYSELEKPLQVYV-DYVELPMQELDW---CGVEATVQRAKLT 104

146 AASGPCNREVLLED-----MOTKKCOLWTFESHALVDVTFQQRVLSRVEAAYKH 195

105 EFLECDNSGNFNLQOPLMRCVIRLEANKYQFINSHHLLMDGWCNGIILLTEVFLYK- 163

196 EKDTHRPETPESSDAT-DTDSQSVSVSMSCEDNAVYSATHFWOTHLNDLNA----- 245

164 -----AFQSDVDTLPSPIPGVNLWLQOQNEQAESEFRKQLQDFTAPTIVLGNKIN 216

246 ---SVFPHLSHLWLVNPT-----TAEHRTFPLSQALSNAICRTALSTILLSRYTH 296

217 NQSSIF---SQQSVNSQOQTEQLQNFVROHRLTL-----NILVQGAWALLLSKYG 264

297 SDEALFCAVTE---QSLPDKHYLADQTYQTVAPLRVHCQSNLRASDVMDAISYDDRLG 353

265 ENNVLFGTIVSGRPPELQVESIV--GLFINTLPVRVDVNGEMQIIPWLQLOQAQVERE 322

354 HLAFFGLDIRNTGNGSACDFQTVLLVTD-GSHVNNINGFLQ--QITESHFMPCNN 410

323 QYSYSLVDITQSSVPRGTSFESLVLFPENPVSTESVIKTFDQTLQIEDTQGFECTNY 382

411 RALLLHCMESSGALLAVYDHNVIDSLQTLRLQFQHLIKLQSPLD--LSSMAEVLN 468

383 PLTLTVIPGEE--LLFKYSDVSYFDAATISRLT---GHLETLMSVNEPEORLSQISM 437

469 MTEYDRAET--ESWNSQPLEVQDQTLIHHEMLKAVSHSPKTAIQAWDGDWTFYSLDNVSSR 527

438 LTPQEKQLIKDWNQTTQAYPNQC-IHQLFEAQVETPDALVLFGECLTYRELNNKANQ 497

528 LAVHIKSLGRAQQAIIPIVFEKSKWVIASMLAVLKSGNAFTLIDPNDPAPTAQVVTOT 587

498 LAHYLQQLGVK-PEILVIGCLERSLEMTVCLLGIKAGGAYVPIDPEYPOERIAVLMEDS 556

588 RATVALT-----SKLHRETQKLVGRVVD--DELLQSVSADDFSSLTKSODLAVIET 641

557 QKVFLTQQLLAQIAPQNTHKI---CIDTEWDKISRQNTNPD--SGVKLONLAVIYIT 611

642 SGSTGDPKGMIEHRAFSKALKFGASGLGINSIDRALQFGTHAFGACLLIEIMTTILINGC 701

Db 612 SGSTGKPKGVNTHQGTICNLLNMQETQYQINSTDSTILOKTPFSPDVSVWEEFTWLLTGAR 671

QY 702 VCIPSDDDRMS--IPSFIRYNNVMMATPSPVMTGTF-----SPEDVPLGLATLVLVGQMS 755

Db 672 LVIAKPGGHRDSAYLINLIIQEQITILHFVPSMLRIEESGNVEKCTSLKRVICSGALS 731

QY 756 SSVNAIWAPKL--QLLNGYQSSSS-SICF-----ASNMTSEPNMGRV-GAHSWIDP 806

Db 732 LDQONQFLERDFCDELYNLYGTEAADIWTFQCRKHTNLKTP--ICRPANTHIYILDQ 789

QY 807 NDINRLVPIGAVGELVIESGIARDYIVPP--PPEKSPFTDIPSW-----YPANTPDG 859

Db 790 H--LQPVPIGVPGEIYIGGAGVARGYLNRPDLTAERFISYSDGSLGILYTPMPHAPCIP 847

QY 860 AKLYRTGLARYASDGSIVCLGRIDSQVKIRQORVELGAIETHLRQOMPDDLTWVEATK 919

Db 848 HALYKTDGRARYLRDGTIEYLRGLDNQVKIRGFRIELGEIETVLSQISEVQAAIVVT-- 905

QY 920 RSOSANSTSLIAFLIGSSYFGNRPSPDAHIIDHDATKAINIKLEQVLPRHSIPSYICMLE 979

Db 906 -EDTTGDKRLVAVIVS-----ENPEITD-----LRQYLAKNLDPYMIPOFITLDA 950

QY 980 LPRTATGKIDRRRLRMKIDLDKQTOGAIVQAPAPIPVF-----ADTAALKHS 1029

Db 951 LPLTPNGKVDRRAL-----PKQFKTSSHTIAPRTPTTEELLN 988

QY 1030 IWYQSLGIDPATYVNGATFFELGNGSITAIKMYNMARSV-GMDLKVSNIIYQHTPLAGISA 1088

Db 989 IWQNILNLN--SVGITDNFFALGSHSLVLRVLSQIQQVFGVDISLRQIFENPTIAELAT 1046

QY 1089 VVKGDPLS-YTLIPKSTHGEVPSQSGRLWFLDQDVGSLWYLIPIYAVRMGRGVNDA 1147

Db 1047 LINSKLSLYSSIPERESTDKILLSFAQRLMLAQIPEPNPSYVNAALQLTGDVNDV 1106

QY 1148 LRALAALAEORHETLRTTFEDQGVQIVHEKLSSEMKVIDLCGSDLDPF-----EV 1200

Db 1107 LTQSOEIIHRHETLRTSFVNDVQGGIATVATEINSHIPVVDL--SPLPSSQOKLVOEL 1164

QY 1201 LNOEQTPFNLSEAGWRATLLRLGEDDHILTIVMHIIISDGWSIDVLRDLNLQYSAAL 1260

Db 1165 AQEESQPPNLEISPLLRKAKFLYNTQEXILLTTHHIIITDGSINVFAQEMATIYQAFS 1224

QY 1261 KDSKDPLSALTPLPIQYDFAKWQKQDFTEQKOLNWKKQKLDSSP--AKIPTDFARPA 1318

Db 1225 QGQSSPLQ---PLKIQYADFVWORSQDKFNQVEYWRQOL--ESAPPELLDLTPDYRPA 1280

QY 1319 LLSGDAGCVHVTIDGELYQSIRACNEHNTTSPWLLAAPRAAHYRLTAVEDAVIGTPIA 1378

Db 1281 IQTFGRQHSFTTISEELTEKIKOLGOETHTLFWVWFSAFSLILHRYSQEKIVIGSPA 1340

QY 1379 NNRNRELDIIGCFVNTQCMRINIDHDTFTGLINQVKATTTAAFNEDIPFERVYSALQ 1438

Db 1341 NRHYPTGELIGFPVNTFALLISLEDNPTVAELLPRVREMLVLSAYSHQDVPEQVVEELK 1400

QY 1439 PGRDLSSTPLAQLIFAVHSQKDLGRFKOGLESVPVPSKAYT-RDMEPHLFQETDS-L 1496

Db 1401 L-VRSLSHSPLFQVMLAVEN-APTQPIETPLGRWSPLEIDGGTAKFEDITL-MAIETDAGL 1457

QY 1497 KGSVNFADLEFKMETVENVRVFEILLRN---GLQSSRT--PVSILPLTQD-----IVTLE 1547

Db 1458 QGKWEYNCDLF---TETTHRETHLQNLILLGWTSEPSQKISDLPLMSGELQLOIIFNG 1513

QY 1548 KLDVLNVKIVDVPRESSLADVFQTVQSVAYSDSLVAVVDSCLRTYITELDQSDILAGLRR 1607

Db 1514 TSPTNLIH-----CIQELFEQVAKFGEIAVICQHQKLTYSLELTKANLAVHLKS 1566

QY 1608 RSPAEITLAVAPRPSCEITVAFVGLKANLAYPLDVRSPSPARVODIISGLSGPTIVLI 1667

Db 1567 LGVQPEITAVGICVHRSLDFIIGLAILKAGGFVPLDPTYPQERLEFLIED-AQIQVILT 1625

QY 1668 GHDTAP--PDIEVTNVFVRIRDALNSNADGFEVIEHDSKTP-----SATSLAYVLYTSG 1721

Db 1626 QOQHIFQPLDPLI-----FCFDDTDAN-----LTQHPTNPVNITTENLAYVMTSG 1673

Db 1830 LPEVNVQYKDYVWHTFNQSAAMKQEA-----YWLVDYVRIPIPSKLDPPDYKRR 1880
 QY 251 -----LSDLHVPNTTAEHRITFPLSOKALNSAICRTALSILLR 293
 Db 1881 HDTFEGSSVFELEMERELSDHI-----RKLAKNGT-----TLTYVMLSAYVLLNK 1927
 QY 294 YTHSDEALFCAVTEQSLPFKHYLAD--GYQTVAPLVRHCQNLRSADYMD-----AIS 346
 Db 1928 YTNQTDIVGTGAAGRLHPD---LQDVFGVFNVLALRNEVDTSYFKEFLQOTKERTIA 1984
 QY 347 SYDRGLGHLAPG-----LRDIRTGDNGSAACDFQTVLLVTDGSHVN-----GINGFL 396
 Db 1985 AFON---SEYPPFDLIRKLNGVRESNRNL---FDTMEVLEDAFMETKQKGVKLSPII 2037
 QY 397 QOITESHENPCN-----NRALLHCOMESSGALLVAYYDHNVDTSLOTTLLQOFGHLIK 452
 Db 2038 FELDNKAFMIFNVLDFEQKIVLINIEYSTS-----LFKDETI-----QKIAEDYFRILE 2086
 QY 453 CLOSPLDLSMAEVLNLTMYD-RAETESWN-SOPLEVQDTLIHHEMLKAVSHSPTKTAIQ 510
 Db 2087 EVSENLDV-ALHOIDMISROEKXTLLESFNHTKTAIYPKGAIHQLEEQAKRIPDHTAVV 2145
 QY 511 AWGDWYSELNVNSSLAVHIKSLGLRAQAAIPIVYFEKSKWVIASMLAVLKSGNAFTL 570
 Db 2146 FEDOKLTYROLNEKANQAVALLREKGVK-PDVLVGIMMERSSDMIAAILGVLKAGGAYLP 2204
 QY 571 IDPNDPARTAQVVTQTR-----ATVALTSKLHRETVQKLVGRVVDDELLOSVASDD 625
 Db 2205 IDEPEYPERMYNAFSEVKVVIISDVPLABELTAESIE-----LIHMDDERIAGODRS-D 2258
 QY 626 FSSLTKSQDLAVYFTSGTDPKGMIEHRAFPSSCALKFGASLGINSIDTRALQFGTHAF 685
 Db 2259 IDNVNOSGDLAVYVTSGSTGPKGVNIEHQSLINLCSWHQSCFVQONDNSIYASISF 2318
 QY 686 GACLLBIMTTLNGCVCIPSDDDRMNSIPSFNRYNVMNMAPTSMYGT-----FSPED 740
 Db 2319 DAFVWELFFYITAGATVHVLNQETRLD--VEKLNRYFHDHII-TISFLTPVCEQFTALD 2375
 QY 741 VPLGLATLVLGVEGSSNAIWAAPKQLLNGVQSESSSIC--FASNMSSTEPNMGRVAG 798
 Db 2376 NISLRTLTLGGDKL-----NVFEKSYOIVNNGTPTNVVATSFPIKSHQNIPIKPID 2431
 QY 799 AHSWIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPEKSPFFTDIPSWYPANTFPD 858
 Db 2432 NVKYYLNLKDL-QCLPLGASGELCIAGELGARGYVNRPELTREKFI-----GNPFVP 2482
 QY 859 GAKLRTGDLARYSDGSIVCLGRDSQVIRGORVELGAIETH-LRQOMPDDLTITVREA 917
 Db 2483 GERMVYTGDLAKMLPDGNTQFLGRVDQVQKIRYRIEPEGEIENRLLKYEKIEEAAVIA-- 2540
 QY 918 TKRSQANSSTLAFILIGSSYFGNRPNSDAHILDHDKATKAINIKLEQVLPRHSTPSYICM 977
 Db 2541 --REDGDHPYLCAYV-----VKVEPEKIRAFLLKSLPDYIMQYFVOL 2585
 QY 978 LELPRTATGIDRRRLRIMGKDI-LDKQTOGAIVOQAPAPIPVFADTAALKHSIWQSIG 1036
 Db 2586 DGLPLTVNGKVBKSLPVPEVSRTVMDR-----YEAPRD-----QMEKLVSIWEEALG 2634
 QY 1037 IDPATVNVGATFELGCGNSITAKWVN-MARSGVMDLKVSNIYQHTPLAGISAV-VKGD 1094
 Db 2635 IN-KIGINSHEFEAGHSLKAAALVSTHKELVNKLPLRQIFETPTIKGLRDISVRRRK 2692
 QY 1095 LSVTLIPKSTHGPVEO-SYSOGRLEWFLDOLDVGS-LWYLIPYAVMRGPNVNDALRRAL 1152
 Db 2693 CFY--IDRTEKPYRSLSSAKRLLISQ--TGSHVAYMFMFAMTLEGDFDIRFENTL 2748
 QY 1153 AALEQRHETLRTTFEDQDGVGVQIVHEKLSSEMKVIDLCSGLDPPEVLNQOTQTPFNLS 1212
 Db 2749 KMWKRHESFRTSFVIMIDVEMQOIKEIDFOVAYSDI--CKESAEEKIKSFIRPFHLE 2805
 QY 1213 SEAGHRATLLRLGEDDHILTIVMHHIISDGSWIDVLRLDNLNOLYSAALKDKSDPLSALT 1272
 Db 2806 KAPLLRAEVVKNREHLLMFDMHHIISDGVSTDIFIQELGALYEG-----KSLKP 2856

QY 1273 LPIQYSDFAKWQDOQFTEQE--KOLNYWKOLKODSSPA-KIPTDFARFALLSGDAGCVHV 1329
 Db 2857 FHIQYKDYAEWENSHARSEELKRQEYWLKTYKGDIPVLDLPIDHKRPLTKSSBGDTVTA 2916
 QY 1330 TIDGELYOSLAFACNEHNTTSFVLLAFAAAYRLTAVEDAVIGTGTPIANRREPELEDI 1389
 Db 2917 AIESEFRKLOHMAKENGVTMYMLLAGYTALLSKYTGQBEDIIVGTTPAAGRNHEDIOH 2976
 QY 1390 GCENVTOCMRINTDHDHFTGTLINQKATTTAAFENEDIPFERVWSALQPCSRDLSSTPL 1449
 Db 2977 GMEVNTLAI RNHPGKKTFRDYLOEVKENTLOAYENQDYFEELEKVN-KRDMARNPL 3035
 QY 1450 AQIFAVHSOKDLGRKFKQGLSVVPVSK-AYTRFDMFHLFOETDSLKGSVNFADFLK 1508
 Db 3036 FDTMLVYHN-TDVKPFPEAGRLSRLEIKRGISKFDITVTASEAADGLRLEVEYSTLFN 3094
 QY 1509 METVENVVRVFFETILRNGLOSSRTPVSILPLTGIVTLEKLDVL-NVKHYD--YPRSSL 1565
 Db 3095 KERMERLSEHILSLE---QAADHPDIAINQIDVLTGERHRLYDFNRTDGVFCREMTI 3151
 QY 1566 ADVFOTQVSAYPDSLAVVSDSSCRLTYTELDQSDILAGWLRRRSMAPETLVAVFAPSCE 1625
 Db 3152 PELFEQAKETPDHPAVAFGDETISYRELNERANSLAFTLRQKGVGPDVITAGILTERSIE 3211
 QY 1626 TIVAFPGVLKANLAYLPDVRSPSARVQDILSGLSGPTIVLIGHDTAPPDIEVTNVFEVR 1685
 Db 3212 MIVGIMGILKAGAYLPIDPAYPOERISYIVKD-SDVSVLCAAGDVDPG--EAYTGDII 3268
 QY 1686 IRDALNDSNADGPEVTEHDSKTSATSLAYVLVTSGSTGRPKGMIEHRYI---IRIVTS 1742
 Db 3269 IDQT---GONDHVENLKH--IKPQ--HLAVIYVTSGSTGAPKGMVIEHHVNNLVHGLNE 3322
 QY 1743 GCIPNYPSETRMAHMATIAFDGASYEITYSALLFGRTLVCVDYMTTIDARALKOVFFEHV 1802
 Db 3323 RIYQHLDAHLNALVAPYIFDASVKQIFAALLFGHTLCIVPRETANDAMSLIEYSKNNI 3382
 QY 1803 NAASHVTSQDVPVLRPRRLSRTLMFF----FLVYTDSTAPDALDAQGLYQ--GVOC- 1854
 Db 3383 N-----VSOMTPAHLNMLAYVDKTELEFVGVGLDALTDPDVG--GLFHRFPNLSCN 3435
 QY 1855 -XNGYGTENGVMSTIYPIIDS-----TESFINGVPIGRALNNSGAYVDPPEQOLVGIVM 1908
 Db 3436 ITNVYGTPECCVDAASHQIESGKVPQTPS----IPGRPLNLTISIVDKELRPLVGIA 3491
 QY 1909 GELVVTGDLARGYSK-ALDENRFVHTVNDQTVKAYRTGDRVRYRIGDGLTEFFGRMD 1967
 Db 3492 GELCIAGEGVARGYVNRPELTAEKFDVHPF-EPGKKMYKTGDLAMW-LPDGQTEFLGRAD 3549
 QY 1968 TOPKIRGNRIESAEIEAALLROSSVDRDAAVLQONEDQAPAILGVVADHDHSENDKQS 2027
 Db 3550 HQVKIRGYRIELGEVEEQLLTHEKIKEAAVIAKQDQNGNSYLCAYTASDKELPAADVRF 3609
 QY 2028 ANQVEGWQDHFECSGMYSDIGEIDPSTIGSDFKWTSMYDCSQIDFDEMHEWLGCTTTLH 2087
 Db 3610 LER-----EMPDYMIPSYF-----VKLDLPR----- 3631
 QY 2088 DNRSGLNVLEIGTGSMLFNLDLSRLSYVGLPSRSAAAFVKNKATESIPSLAGAKVQV 2147
 Db 3632 -----TPSGKV-----DRSA-----LPEADGNVNVME 3653
 QY 2148 GTATDIGQVDDLHPDLVNLNSVIQYFP-----SSEYLAETADTLHLHPNVQRIFF 2197
 Db 3654 GTGYD-----PPRNEIERKLVQVNRILGAEDIGISHHFFAAGGDSKALQIVSRL-- 3704
 QY 2198 GQVRQOATNEHFLAARAIHTLGNKATKDDVRKMAELEDMEBELL-VEFAFTSLKDRFP 2256
 Db 3705 AKMNLKLEMKALFANPKIKDLSRFITEETPRHKNKPVGTETELLPIQKRYFANKEELD 3764
 QY 2257 GLUHEVHEILPKMNEAVNEL-SAYRYAAVHVHVGSLGDELVLVPEKDDWDIDFOANLQKS 2315
 Db 3765 HFNQSFMLFRKQDGYDENIVRTAFNKILEQH-----DALRMVIEEKDGLIIOYNRGYREN 3818

| | | | |
|----|------|---|------|
| Qy | 2316 | LGOLLKSSDAIMAVSKIPEITAFERQVVVUNSNIDEMQJSTIRSSAGDSSLSVPDI | 2376 |
| Db | 3819 | LFDL-----DV-----YDRFGFSQ-----EBKVPELATGIOKXSI----- | 3850 |
| Qy | 2376 | FRIAGEAGPRVEYSSARQWSONGALDAVHHCCSQGRTLVNFTDHHLRGSDLLTNRPLO | 2435 |
| Db | 3851 | -----RKGKLVHIGIFRAEGDHLLIAI-HHLVVDG----- | 3880 |
| Qy | 2436 | RLQNRRTAIEVRERLRSLLPSYMPISNIVVLDKMLNANGKVDH-KELSRRAKVVPKQO- | 2493 |
| Db | 3881 | --VSWRLPFDEFTL-----YLOALKEPLDIGYKTDYSOFARQLKYYAQR | 3927 |
| Qy | 2494 | -----TAAPLPTPPISEVEVILCE-EATEVFGMKVDITDHF-----FNLGGH | 2534 |
| Db | 3928 | LLKEREYWKALEADVPFPAEKLELDTEHSAATLSIRIGCPDVTAKLLRANAFKAYTEIN | 3987 |
| Qy | 2535 | SLLATKLISRIDORLKVRTVKDVEDHPFADLASVIROGLGLOQPVSOGODRSAHMA | 2594 |
| Db | 3988 | DILLITALIA-----AVRDITGE-----NKLKVMMEGHG-REDILGV----- | 4023 |
| Qy | 2595 | PRTEETAILCDEFAKVLGFQVGITDNFFDLG-----GHSIMATKLAVR-----IGH- | 2640 |
| Db | 4024 | -----DITRTIGWFTTVYPVIDLGEKEIJSQIKMKVKEALRKIPNKGIGYGV | 4072 |
| Qy | 2641 | RLDITTVSVKDVDPHVPFLQALIALDNLVQSKT-----NEUVGR----- | 2679 |
| Db | 4073 | KYMTTELQKTQOAPLSFNYPGEMNDMRKVPFSQSPFSGESIGGKIVRHCAIEMNAIS | 4132 |
| Qy | 2680 | -----EMAEYSPF-----QL-LFTEDPEFM-----ASEIKPO-----LE | 2708 |
| Db | 4133 | LNGLITYTTFNQDQVOTSTIEGLNOSFRLENLEKIVDHCVDKESDMTFSYGDVSLGLE | 4192 |
| Qy | 2709 | LOEIIOD-----IYPTQMKAFLE-----DHTTARPRPVPEYIDFPSTSEPDAG | 2755 |
| Db | 4193 | ELELIKDKYSAFOIEKTYPLANMKQMLFHNAMDQTSG--AYFQIVIKLGRVHFDI-- | 4248 |
| Qy | 2756 | LIXACESLVNHLIDIFRTVFA-EASGELYQVVLSCDLDPQIVITENINTATNFEFLDEFA | 2814 |
| Db | 4249 | LEESFHEIKVRHILRASFEYEITAPRQIIARDRKTPTSIDLTGENRTRHRTIETVL | 4308 |
| Qy | 2815 | KEPVRLG-----HPLRFTTIKOT-KSMRVIMRISHALVDGLSLEHVVRKLHMLY----- | 2863 |
| Db | 4309 | KEOQKGFDLSSSALMRVCLIKMSDESRYLWISHHHLLDGCWLGIVLSFLSFLYKIMK | 4368 |
| Qy | 2864 | -NGRSLPLPHQFSRYMQY-TADGRESGHGFWRDVION-----TPMTILSD | 2907 |
| Db | 4369 | GESRRUKEPKPYDGIYKLEKQDOBEAVAYWKDYLKGYESRSELPAFNRGATSEYCGRE | 4428 |
| Qy | 2908 | TVVD-GNDATCKALHLSKVINPISOVLRGSSNIITQATVPNAACALVLSRESKDVVFG | 2966 |
| Db | 4429 | KVIFSKELTTKTRIAX-----QHHVTINTVLQGTWGMILAKYKNTDEVFG | 4476 |
| Qy | 2967 | RIYSGOGLPVE-YQDITVGPCITNAVVPRAHIESS-DYNQLLHDIDQOYLLSLPHEITGFS | 3024 |
| Db | 4477 | TVVSGGREA-PVDGIEEMVGLFIHTIPRISFEGARSFKEVLKKTQAESIESNRYSMNLS | 4535 |
| Qy | 3025 | DLKRNCITDWPPEATINFNSCCITYHNFPYHPSQEQORVEMGVLTKVFNTIEMDPLYLAI | 3084 |
| Db | 4536 | EIQVLSMKRELITH--VMAFQNYAF-DEELFRSQSGETGFELEGVH-GKERTNYPNL | 4590 |
| Qy | 3085 | AGEVEDPDGAGLV-----TVIAKTQLFGRRKRVHEHLLLEEVSKTFEGUN | 3126 |
| Db | 4591 | TGVLDEQLKLUKLTFFNENVYDNIITLEKHIIITVAEQAEDFOTLROIN | 4641 |

A-Title: Characterization of the syringomycin synthetase gene cluster. A link between
A-Reference number: Z18153; MUID:99047670; PMID:9830033
A-Accession: T14593
A-Status: preliminary; translated from GB/EMBL/DBJ
A-Molecule type: DNA
A-Residues: 1-9376 <GUP>
A-Cross-references: EMBL:AF047828; NID:g3510628; PID:g3510629; PIDN:AAC80285.1
C-Genetics:
A-Gene: syrE
C-Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C-Keywords: carrier protein; phosphopantetheine; phosphoprotein
F.441-896/Domain: acetate-CoA ligase homology <ACL1>
F.914-981/Domain: acyl carrier protein homology <ACP1>
F.1529-1984/Domain: acetate-CoA ligase homology <ACL2>
F.2002-2069/Domain: acyl carrier protein homology <ACP2>
F.2613-3071/Domain: acetate-CoA ligase homology <ACL3>
F.3089-3156/Domain: acyl carrier protein homology <ACP3>
F.3700-4158/Domain: acetate-CoA ligase homology <ACL4>
F.4176-4244/Domain: acyl carrier protein homology <ACP4>
F.4768-5224/Domain: acetate-CoA ligase homology <ACL5>
F.5242-5310/Domain: acyl carrier protein homology <ACP5>
F.5834-6280/Domain: acetate-CoA ligase homology <ACL6>
F.6298-6366/Domain: acyl carrier protein homology <ACP6>
F.6892-7352/Domain: acetate-CoA ligase homology <ACL7>
F.7370-7437/Domain: acyl carrier protein homology <ACP7>
F.7990-8440/Domain: acetate-CoA ligase homology <ACL8>
F.8458-8526/Domain: acyl carrier protein homology <ACP8>
F.9015-9083/Domain: acyl carrier protein homology <ACP9>
F.946, 2034, 3121, 4208, 5374, 6330/Binding site: phosphopantetheine (Ser) (covalent) sta

F:9015 5083/Domain: acyl carrier protein homolog (ACP)
F:946,2034,3121,4208,5274,6330/Binding site: phosphopantetheine (Ser) (covalent) #sta

Db 2123 PLPEVAGETLAYVIYTSSTGQPKGVAVSQAAALVAHQAAARTYGVGPGDQLOQFASISF 2182
Qy 686 GACILEITWTLINGCVCIPSDDDRMNS--IPSFINRYNNMNMATESYMGTSPE---- 739
Db 2183 DAAAEQLFVPLLAGARVLL--GDAGOWSAQHLADEVERHAVITILDPLPAYLQQAELRHA 2241
Qy 740 -DYPGLATLVLGEOMSSSSNAWAPKQL--LNGYQOSSESSICFASNMTPEPN---MG 794
Db 2242 GRTIARACILGGEAWDASLLTQOAVQAEAFNAYGTEAVITPPLAWHCRTQEGGAPAG 2301
Qy 795 RAVGASHWIDPNDINRLPIGAVGELVIESPGIARDYIVPPPEPEKSPFFTDIPSWYPAN 854
Db 2302 RALGARRACILDAALQCAP--GMIGELYIGQCLARCYLGRPGGTARFVADPFS----- 2355
Qy 855 TFPDGAKLYRTGLARVADSGVLCGRIDSQVKIRQORVELGAIEHLRQOMPDDITIV 914
Db 2356 --GSGERLYRTGLARYRVDQVEYLKQADQKIRGFRIEIGIESQLLAH-----PYV 2408
Qy 915 VEATKRS--QSANSTSLIAFLIGSSYFCNRPDSAHILDHDATKAINIKLEQVLPRHSIPSF 973
Db 2409 AEAVALDGVGPLLAAYLVG-----RDA--MRGEDLLAELRTWLGLRLPAYMQPTA 2459
Qy 974 YICMLELPTATGKIDRRIRIMGKOILDQTOCAIVQOAPAPTPVFPADTAAKLHSIWQ 1033
Db 2460 WQVLSLPLNANGKLDKAL-------PKVDAARRQAGEPPREGLEERS--VAALWEA 2508
Qy 1034 SLGIDPATVNVGATFFELGNSGTAIKHVNAR--SVGMOLKVSNIYOHPTLAGISAVVGK 1092
Db 2509 LLGVE--GIARDEHFFELGHSLSATRVVSRRLQDLDELVDPLRLFRPVLADFAASLES 2566
Qy 1093 DPLS----YTLIPKSTHEGPEVQSYSGRLWELDOLDVGSUWLYLIPYAVRMGRPVNDAL 1148
Db 2567 QAASAPVLQILPR--VAELPL--SHAQRMWFLWKLEPESAAVHLPSVLHVRGVLDQAAL 2623
Qy 1149 RRLAALAEORHETLRTTFEDQDGVGVQIVHEKSEEMKVVIDLGC--SPLDPPFEVLNOBQTT 1207
Db 2624 QOAFDWLVLRHETLRTFEEDVGOART ILANPLRLVLEDCAGASEATLQRVAEEIRQ 2683
Qy 1208 PFWLSSEAGWRATLLRLGEDDHILTIVMHIIISDGSIDVLRDLNOLYSAALKDSKDLPL 1267
Db 2684 PFDLARGPLLRVLLALAGQEHVILVITQHIVISDGSQMVVMDQLLQVAAAARREGQ--- 2740
Qy 1268 SALTPLEIYSDFAKWQKQFIOE--KOLNYWKKOLKDDSSA--KIPTDFARALLSGDA 1324
Db 2741 PTLAPTLQYADYAAHRAWLDSGEARQLDYWRERLGABQPVLEPADRVRAQASGRG 2800
Qy 1325 GCYHVVTIDGXYQSLRAFNCNEHTTSFVLLAAAFRAAHYRLTAVEDAVIGTPIANRNPE 1384
Db 2801 QRLDMALPVSLSEELLACARREGVTPFEMILLASFQVLLKRYSGSDIRVGPVIANRRAE 2860
Qy 1385 LEDIGCFVNTQCMRINIDHDTFTGLINOVKATTTAAFNEDIPERVVSALQPGSRDL 1444
Db 2861 VERLIGFFVNTQVLRQCVDAGLAFRLDGLRGVREAAALGAQHQLPPEQLVDALQP--ERNL 2919
Qy 1445 SSTPLAQLFVAVHSQDKLGRKFQGLSEVPVPSKAY----TREDMEHFLFOETDSLKGSV 1500
Db 2920 SHSPLFQVXNHQS-----GERQDAQVDGLHIESFADWGAAQDFDALDWTWETPDGLGAAL 2975
Qy 1501 NFADELFRKMETVENVVYRFFELIRNGLSRTPVSILPLTDGIVTLLEKLDVLRNVKHDYVP 1560
Db 2976 TYATDLFEARTVERMARHQNLLRGMLENPQASVDSLPLMDAERQQLLEGWNATAEYF 3035
Qy 1561 RESSLADVQTVQVSAYPDLSLVVDSRCTYTYELDRQSDILAGLWRRRSMPTLTVAVFA 1620
Db 3036 LQGVHRLFEQVETPTAPALAFGEERLDYALNRRANRLAHALIERGVGVADRLVGAM 3095
Qy 1621 PRSCETIVAFGVLKANLAYPLDVRSPSARVQDILSGLSGPTIVLIGHDTAPPDIEVTN 1680
Db 3096 ERSIEMVVALMAILKAGGAYVPDPEYPERQAYMLEDSCVQLLLSQSHKLPLAOGVOR 3155
Qy 1681 VEFVIRDALNDSADGFEVIEHDSKPSATSLAYVLYTSGTGRPKGVMIHRRVIRTV 1740
Db 3156 IDLDQ--ADAWLENHAENNPICIE-----LNGENLAYVIYTSGSTGKPKGAGNRHSAL---- 3205

Qy 1741 TSGCIPNYPSETRMAHM-----ATIAFDGASYEYISALLFGRTLVCVDYM 1785
Db 3206 -----SNRLCWMQOAYGLGVGTVLQKTPFSEDSVWEEFWPLMSGARLVAAAPG 3255
Qy 1786 TTLARALKOVFFREHVNAAASHVTS-----SSODVPLVRPRLRSLRSLTFFLFFLVVTDSTA 1839
Db 3256 DHRDPAKVALINREGVDTLHFVPSMLQAFLOQDEDV-----VSCSLKRVCSGEALS 3308
Qy 1840 PDA-----LDAQGLYQGVQCYNGYGTENGVMSTIYPIIDSTESFINGVPIGRALNNSG 1892
Db 3309 ADAQQQVFAKLPOAGL-----YNLGYPTAAIDVTHW--SCVEEGKDAVPIGRPIANLG 3360
Qy 1893 AYYVDPBQQLVGIVGMGELVVTGDLARGYSDK--ALDENRFV--HITVNTQYKAYRTGD 1949
Db 3361 CYILDGDLPEVPVGVGLGELVLAGRLARGYHQRPGLTAERFVAFSPVAGE---RMYRTGD 3417
Qy 1950 RVRYRGDGLIEFFRGMDOFKIRGNRIESAEIETAALLROSSVDAAVVLOQNEQDAPEI 2009
Db 3418 LARYR--ADGVIEYAGRIDHVKRLGRLEIETEARLLEHPWVREAAVLAVDSR---QL 3472
Qy 2010 LGFWADHDHSENDKQOSANQVEGWODHFEFGMY-----SDIGEIDPSTIGSDFKGW 2061
Db 3473 VGYVLESECGDWRREALAHLATSLPEYVMPAOWLALERMLPSLNGKLDKRALPRPAAA 3532
Qy 2062 TSMYDGSQIDFDEMHEWLGETTRTLHDNRSLGNVLEIGTSGMILFNLDSRLESYVGLPE 2121
Db 3533 GQTHVAPQNDMERRIAAAMVADVLKLEEVGATDNFFALG--GDSIVSIQVVSRCRA--AGIQ-- 3589
Qy 2122 SRSAAFVYNKATESIPSLACKAKV--QVGTATDIGOVDLHDPDLVLSVTOYFPSSYLA 2180
Db 3590 -----FTPKDLFOQQTVOGLARVARVGAQVMEQ-----G 3619
Qy 2181 EIADTLILHPNQVRIFFGVDVRSQATNEHFLLAARAIHTLGNKATKDDVRQKMAELEDMEEE 2240
Db 3620 PVSGETVLLP--FORLFF-----EQIPNQRHWNQS 3648
Qy 2241 LLVEP---AFTTSLKORFPLGVEHVELPKNMEAVNELSAIRYAAVHVHVGSLGDELVLPL 2297
Db 3649 LLLKPREALNAKALEAALQALVEHIDALRLRF---HETDGTWHAEE--HAETLGLALLWR 3703
Qy 2298 VEKDDMIDFOANL---NOKSL---GDLKS-----SDAAIMVASKIPEITAFERQV 2344
Db 3704 AEA---VDROLSEICEQSORSLODLADGPLRLSLVDMDAGG-----QRL 3746
Qy 2345 VASLNSNID--EWQL-----STIRSSAEGDSSLSVPDIPRIAGEAGFREVSVSAROWSQ 2396
Db 3747 LVITHLVVDGVSVRILLEDLQRAYQOSLGEAP-----RLPG-----KTSFFKAWA-- 3792
Qy 2397 NGALDAVFHCCSQSGRTLNVNPTDHLHRS-----DLTNRPL-----QRLQ 2439
Db 3793 -----GR-----VSEHARGESMKAQLQFWRELLEGAPAEPCBHPGCALE 3832
Qy 2440 RRIAIEVREEL--RSLLPYSMPIPSNIVLDKMPNLNANGKVDKRLSRAKVVVQQOATAAPL 2498
Db 3833 QREATSVQSRFDRSLTER-----LLQOAPAYRTQVNDLLLTALARYVVCWMSGAS-- 3882
Qy 2499 PTFPISEVEVILCEEATEVFGMKVDITDHFNLGGHSLLATKLISR-----IDQLRKVRIT 2554
Db 3883 -----SSLVQLEGH-----GREELFADIDLSTVGTWFTSLFPVRLS 3918
Qy 2555 VKDFDHPVADIASVIRQGLGQQPVSD--GQCDRSASMAPRTETETAILCDFAKVLGF 2613
Db 3919 -----PV--ADLGESLKAIKEQLRAIPDKGLGYLLRLVLAG-----EESARVLG 3961
Qy 2614 --QVGTIDNFFDLGGHSLMATKLAV-----RIGHRLDTTVSVKVDVFDHPVLFQLAIALD 2665
Db 3962 LQOARITFNY--LQGFDAQFDEMALDIPAGESAGAEMDP-----GAPLD 4003
Qy 2666 NLVQSKTNEIVGREGMAEYSPFOLLETPDEPEEFMASEIKPOL----- 2707
Db 4004 NML--SLNGRVFDELSDWSFSQMEGEOVRRLADDDYVAELTALVDFFCDSPRHGATPS 4062

QY 2708 -----ELQEIIDIYESTQKAFLEFHTTARPRPEVFIYDFPSTSE 2750
Db 4063 DFPLAGLDQARLDALPVALBE-VEDIYPLSPMQQMLF-HSL-----YEQASSDYINOMR 4115
QY 2751 PDAAGL-----IKACESLVNHLDFRTVFAASGELIQVVLSCLDLPQVIE-----T 2798
Db 4116 VDVSGLDLPRFRAAWQSALDRHALRSQFA-WQELQQ-----PLQIYVQRQLPFA 4166
QY 2799 EDNINTAINE---FLDEFAPKPV---LGHPLRFTLIKQTK-SMRVIMRISHALYDGL 2850
Db 4167 EEDLSQANRDAALLAAAEERGFELQRAPLRLLLVKTAEHHLIYTHHLLDQW 4226
QY 2851 SLEHVVRKLMYNGRSLLPHQF--SRMYQYTA-----DGRESHGFRWDVIONTPMIL 2904
Db 4227 SNAQLLESVSYAGRS---PEQLRDRGRYSYIAWLQRQDAAAEAFWRE-----QMAAL 4278
QY 2905 SDDT-VVDG-----NDATKALHLSKIVNIPSVQLR--GSSNIITQATVFNACALVLS 2955
Db 4279 DEPTRLVLAQPLGTSANGVEHLREVDATATARKLDFARHQVTLNLTVOAGWALLLQ 4338
QY 2956 RESDKDVVGRIVSGRG-LP-VEYQDIVGPTNAVPVRAHI-ESSDYNOLLHDIQDQY 3012
Db 4339 RYTQHTVVFGATVSGRPADLPVENQ--VGLFINTLPVVVTLAPQMTLDELLOGLQRQ 4396
QY 3013 LLSLPHTIGFSDLRNCTDPEAITNPFSCCIYHN-----FEYHPSESQEQQVEMGYLTK 3069
Db 4397 LALREQHTPLFELQR-----W-----AGFGGEAVFDNLVFNYPVDEVLERSAGGV--R 4446
QY 3070 FVNTEMDEPL-YDLAIA 3085
Db 4447 FGAVAMHEQTNYPLALA 4463

RESULT 13

T31075
tyrocidine synthetase 2 - Brevibacillus brevis
C:Species: Brevibacillus brevis
C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C:Accession: T31075
J: Mootz, H.D.; Marahiel, M.A.
J: Bacteriol. 179, 6843-6850, 1997
A:Title: The tyrocidine biosynthesis operon of Bacillus brevis: Complete nucleotide sequence
A:Reference number: 220969; MUID:98012987; PMID:9352938
A:Accession: T31075
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3587 <MOO>
A:Cross-references: EMBL:AF004835; NID:g2623770; PID:g2623772; PIDN:AAC45929.1
C:Genetics:
A:Gene: tycB
C:Function:
A:Pathway: tyrocidine biosynthesis
C:Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:513-957/Domain: acetate-CoA ligase homology <ACLI>
F:973-1043/Domain: acyl carrier protein homology <ACP1>
F:1553-1992/Domain: acetate-CoA ligase homology <ACLI2>
F:2010-2078/Domain: acyl carrier protein homology <ACP2>
F:2590-3025/Domain: acetate-CoA ligase homology <ACLI3>
F:3043-3110/Domain: acyl carrier protein homology <ACP3>
F:1007,2042,3075/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 11.3%; Score 1816; DB 2; Length 3587;
Best Local Similarity 22.0%; Pred. No. 5.3e-97;
Matches 844; Conservative 577; Mismatches 1359; Indels 1048; Gaps 140;

QY 31 SYEQFLHYGLDSSRIEAIKCTPPQLMDICNALDKQ--SAIGHAVYDVPTDIDISRA 88
Db 5 SKEQVQDMYAL-----TPMQEGMLFHALLDQEHNSHLVQMSISLQGLDVLGFLT 53
QY 89 LAWKIEIVNOTPALRAFAFSDSGKTSOVLKDSFV-FSMCMWSSSSSPDEVVDRERAAAA 147
Db 54 DSLHVLVERVDYVFTFLFLYELKQLPQLQVVLKQRPPIEFYDLSDACDESEKQLRYTQYKRA 113

QY 148 SGPRCNR-----VLLEDQTKKQCLVWTFSHALVDVTFQORVLSRVFAAYKHEKDTH 200
Db 114 DOERTFLHAKDLMRVALFQMSQHDYQVWSPHHILMDGWCFSIIFDLDLAIYLSQN-- 171
QY 201 RPEPESDSDTDTDSQSVSVSMSCEDNAVSAWETHWQTHLNDLNASVPHLSHLWVNP 260
Db 172 --KTALSLEPVQPSRFINWLE---KONKQAAINW-----SDYLEAVEBQ 211
QY 261 TTT-----AEHRITFP-----LSOKALSNSAICRTALSILLSRYTH 296
Db 212 KTTLPKKEAFAKAFQPTQYRFSLNRTLTQKGTATSONQVTLSTVQIOWGLLQKNA 271
QY 297 SDEALFGAVTE---QSLPFQKHYLADGTYQTVAPLRHCQSNLRASQVMDAISSYDRL 352
Db 272 AHDVLFSGVSVGRPTDIVGDKW--GLFINTIPFRVQAKAGQTFSELLQAVHK----- 323
QY 353 GHLAPFGLRDRNTGONGSAACDFQTV-----LLVTOGSHVNNGIN-----G 394
Db 324 -----RTLQSQPYEHVPLYDIQTSVYLKQELIDHLLVIENYPLVEALQKALNQIG 375
QY 395 FLQQTESSEHMPCCNNRALLLHCQMESSGALLVAYYDHNVIDSLQTLRLQOF-GHLIKC 453
Db 376 F--TTTAVEMEFTNLTVMVPKEE-----LAFREDYNA--ALFDEQVYVQKLAGHLQOI 427
QY 454 LQSPDLDS--SMAEYNLMTYEDRAIESNWSQPLEVQDPTLIHHEML-KAVSHSPKTAIQ 510
Db 428 ADCVANNSGVELQCIPLLTEAETSQLLAKRTETAADYPAATMHELFSRQAEKTPQVAV 487
QY 511 AWDGWTYSELDNVSRRLAVHIKSLGRQAQAIIPVYFEKSKWVTASMLAVLKSGNAFTL 570
Db 488 FADQHLTYYRELDEKSNQLARFLRKGI-GTGSVLGTLLDRSLDMIVGILVLKAGAFVP 546
QY 571 IDPNPPARTAOVVTQTRATVALTSKLHRETV---OKLVGRCVVVDDLLQSVASDDF 626
Db 547 IDPELPAERIAVMTLSRVPLVVTQNHLEAKVTTTETIDIINTAVIGEE-----SRAP 600
QY 627 SSLTSKSDLAYVFTSGTGPDKPMIEHRAFSSCALKFGASGLNSDTRALQFGTHAFG 686
Db 601 ESLNQPHDLFYIYITSGTGPQKVMLEHNRNANLHMFTFDQTNIAFHEKVLQYTCSPD 660
QY 687 ACCLIMTTLINGGCVCIPSDDDR--MNSIPSFIRYNNVNMWMTPTSYNG-TFSPEDVPG 743
Db 661 VCYQEIFSTLLSGGQLYLITNELRRHVEKLFQIEQKISILSPVSKLFIFNEQDQAO 720
QY 744 -----LATLVLVGEOMSSSVNAIAPKLO-----LLNGYGOSESSSICFASNMSTE 789
Db 721 SFPRCVKHLIITAGEQL-----VVTHELQKYLQRHVRVFLNHYGSETHVV---TTCMD 771
QY 790 PNN-----MGRV-GAHSWVIDPNDINRLVPIGAVGELVIESPGIARDYIVPPPEKS 841
Db 772 PQGAPELPPIGKPISTNTGIYILDEG--LQLKPEGIVGELYISGANVGRGLHQPELTAE 829
QY 842 PFTDIPSWYPANTPDGAKLYRTGDLARYASDGSIVCLGRIDSQVKIRGORVELCAIET 901
Db 830 KFLD-----NYPQGERMYRTGDLALWLPDQGLEFLGRIDHQQVKIRGHIEGEIS 881
QY 902 HLURQMPDPLTIVWEATKRSQSANSTSLIAFLIGSSYFCGNRPSDAHILDHATKAINKL 961
Db 882 RLNLHPAIEAVVID---RADETGGKFLCAYVVLQKALSDEEMRAY-----L 925
QY 962 EQVLPRHSPIYICMLEPRTATGKIDRRRLRMKGDILDKOTOGAIVQQAPAPVPA 1021
Db 926 AQALPEYMIPIPSFVTLERIPVTPNGKTDRLA-----PKPEGSAKTKADYVAPT-T 975
QY 1022 DTAAKLHSLWQSLGIDPATVNVGATFFELGNSITAIKWN-MARSGMDLKVSNYQH 1080
Db 976 ELEQKLVAIWEQILGVSP--IGIQHFTFLGHSKAIQILISRIQKECADPLRVLFQ 1033
QY 1081 PTLAGISAVVK-GDPLSYTLIPKSTHEGVPQESYSGQRLWFLDQLDGLSVLWLPVAVRM 1139
Db 1034 PTIQALAAIVEGEEESAYLAIQAEQPAIYVSSAKRMLILNQLDPHSTVYNLPVAMIL 1093

1140 RGPVNDALRRALAEORHETLRTTFEDODGVGVQIVHEK-----LSEMKVIDLC 1191
1094 EGTLEKARLEHATSINLSRESRTSFTINGEPVRSIRHEQGLHPVIVYETAEEQVNEVI 1153
1192 GSDLPPEVLNQBTTPFNLSSEAGNRATLLRIGEDDHILTIVMHHIISDGWSIDLRRD 1251
1154 LGFMQPFDLV---TAPL-----CRVLGVLKLAENRHVLIIDMHHIISDGVSQILIND 1202
1252 LNOLYSAALKSDKPLSALTPLPIQYSDFAKWK--DQETEQKQLNWKKOLKSDSPA- 1308
1203 FSRLYQN-----KALPEQRHVKYKFAVWEKAWTQTDYQOEKYWLDRFAGEIPVL 1253
1309 KIPTDFARALLSGDAGCVHVTIDGELYOSLRAFCEHNTTSFVLLAARAHRYLTAV 1368
1254 NLPMDYPRPAVQFEGERYLFRTEKQLLESLODVAQKTGTTLYMVLAAHYVLLSKYSQ 1313
1369 EDIVIGTPTANRPELIDIGCFVNTQCMRINIDHDTFGTLINOVKATTTAAFNEDI 1428
1314 DVMIGTVTAGRVHPDTEMTGNFVNTIAMRQSAPTFRQFLLEVKDNTLAAFEGQY 1373
1429 PFERVVSALQPSRDLSSSTPLAQLIFAVHSQKDLGRFKQGLSVP-VPSKAYTRFDMF 1487
1374 PFEELVEKLA1-QRNRSNPLFDTLFLQNM-DADLIELDGLTVTPVPEGEVAKFDLSL 1431
1488 HLFQETDSLKGSYNFADELFKMETVENVRVFEILNGLQSSRTPVSLPLTDGIVTLE 1547
1432 EASENOAGLSFCFECTKLFARETIERNSLHYLIQILO--AVSANTQELAQI-EMLTAE 1488
1548 KLDVL---NVKHVDYPPRESSLADVFOTQVSAVPSDLAVDSSCRLYTELDROSIDLAW 1604
1489 KOELLVHFNDAALYPAESTLSOLFEDQAQKPEQTAVVFGKRLIYRELNERANQLAHT 1548
1605 LRRSMPAETLVAVFAPRSCETTIVAFVGLKANLAYLPDLVRSRPAVODIISGLSGPTI 1664
1549 LRAKGVAQSGVIMAQRSLEMAIGITAILKAGGAYVPIDPDYPNERIAYMLEDCR--RL 1606
1665 VLIHGHTAPDIEVTNVEFVRIRDALNDSNADGFEVIEHDSKP--SATSLAVLVTSGS 1722
1607 VLFQOOLA--EKTANVECLYL-----DEEGSYSPQENIEPIHTAADLAVIITYSGT 1657
1723 TGRPKGMIEHVRVIRTVT-----SCIPNYPSETRMAHMTIAFDGASYEYI 1770
1658 TGRPKGMVHEHGVINSVTVNRDEFALSVRDSOTL-----SLSFADFALTFF 1706
1771 SALLFORTLCVDYMTTLDARALKVDFREHVNAAHSVTSQSDVPLRVPRLSRLTMFF 1830
1707 TLIVSGSTVYVLMPEHAKDPIALRNLIJAWECSYVVFVPS-----MF 1748
1831 FLVVTSTAPDALDAOGLXOG-----VQCYNGYGPTEGYMSTIYPI 1872
1749 QALCESTPADIRSIQAVMLGGEKLSPKLVOLCKAMHPQMSVMNATGPTESSVMAT-YLR 1807
1873 DSTESFINGVPIGRALNNSGAYVVDPEQOLGVGMGELVVTGDGLARGYSK-ALDENR 1931
1808 DTQPD--QPITIGRPANTAIYVDQHOLLPLPVGVGEICIGHGLARGYWKPELTAEK 1865
1932 FVHITVNDOT--VKAYRTGDRVRYRGDGLIEFFGRMDTQFKIRGNRIESAETEAALLR 1989
1866 FV---ANPAVPGEMRYKTGDLGRW-LHDGTIDFGRVDQIKVRYGRIEVEGEIEAVLLAY 1921
1990 SSVRDAAVVLOQNE----- 2003
1922 DOTNEAIVVAYQDRGDSYLAAYVTCKTAIESELRAHLLRELPAWVWPTYLIQLDAFPL 1981
2004 -----DOAPEILGFVVADH----- 2017
1982 TPNKGVDRKALPKPECKPATGAAYVAPATEVEAKLVAIWENALGISGVGLDHFPELGGH 2041
2018 -----DISENDKGS----- 2027
2042 SLKAMTVVAQVHREFQIDLLLKOFFAAPTIRDLARLIEHSEQAAGAAIQAPAEQAYPYPS 2101
2028 -----ANQVBGWQDHFEES-GMYSIDIGEIDPSTIGSDFKGTSMYDSQIDFDEMHEW 2078

2102 SAQORMYLLHOLEGAGISYNTPGIIMLEGKLDREQLANALQALVDRHDIURTSFEMVGD- 2160
2079 LGETTTLHDNRSLGNVLEIGTSGMILFNLDLSRLESYVGLPSR---SAAAFVNRKATES 2135
2161 --ELVKQIHDR-----YAVNME-----YVTAEEQOIDDLEFAHVRPDL 2198
2136 IPSLAKAKVQVG-----TATDIGQVDDLHPDLVL-----NSVIQY----- 2172
2199 VPPLLRMSLVKLADERHLLYDMHHTAADAASITILFDELAELYQGRELPMEIRIQYKDF 2258
2173 -----FSSSE--YLAETADTLIHL-----PNVQRIFFGCDVRSQATNE----- 2207
2259 VMOKALHESDAFKQOEAYWLSTFAGNITAVDPTDFRPAVKFAGQVTLSDMQDELLSA 2318
2208 --HFLAARAHHT-----LGNATKDDV-----ROKMAE----- 2233
2319 LHLELAHTNTLPMVLLAAYNVLLAKYAGQDDLIIVGTPISGRSRAELAPVVGMEVHTLAI 2378
2234 -----LEMEELLVPEAF-----FTSLKD-----RFPG-----LVE 2260
2379 RNKPTAEKTFKQLOEVKNAL--DAFDHODYPFESLVEKLGIPRDPGRNPLFDTMFIQ 2436
2261 HVEILPKNM-----EAVNELSAYRYAAVHVHRSGLD----- 2292
2437 NDELHAKTLDQLVRYPYESDSALDVAKFDLSFHLTERETDLFLREYCTKLFKQOVTVERM 2496
2293 -----ELVLPVEK-----DWIDFQANQLNQ-----KSLGDL 2319
2497 AHHFLOILRAVTANPENELQEIEMLTAAEQMLLVAFNDHREYRADOTIQQOLFEELAEK 2556
2320 LKSSDAIIMAVSKIPP-----EITAFER-----QVVASLNSNIDEMQULSTRSSA 2364
2557 MPEHTALVFEEKRMSFRELNERANQAAVLRKRGVGAQIIVALLVERSAEMVATLATLK 2616
2365 EGDSSLSV-PDI--FRITAGEAGRVEVSSAROWSONGALDAVPHCCSOGRTI- 2414
2617 AGGAFLPVDPDYPERIR---YMLEDSQAKLVVTHAHL---LHKVSSQSEVVVDVDDPGS 2669
2415 -----VNPPTD-----HHL----- 2423
2670 YATQTNLPCANTPSDLAYIITYSGTGKPGVMEHKGVANQAVFAHLGVTPQDRAG 2729
2424 -----RGSDDLNRPLQRLN-----BRIATEVRELRSL-PSYMI- 2459
2730 HFASISFSDASVDMFGPLLSGATLYVLSRDVINDFQFAEYVRDANAITFLTPPTAIYL 2789
2460 -----PSNIVLVDKMP-----LNANGKVDKRLSRRAKVVVQKQTAAPL 2498
2790 EPEQVPSRLTLITAGSASSVALVDKWKKEKTYVYNGYGPTESTVCATILKAKPDE---PV 2845
2499 PTF-----PISEVEVILCEEATEV-----FGMKVDITDHF- 2531
2846 ETITICKPIONTKLYIVDQOLQKAPQMGELCISGLSLARGYWNRPETAEKFDVNDPVF 2905
2532 GGHSLAT-----KLISRIDQRLKVR-----ITVKDVF----- 2559
2906 PGTKMYRTGDLARWLPDGTIEYLGRIDHGVKIRGHRVELGEVSVLLRYTVKEAAAIH 2965
2560 --DHPVFADLASVIROGLGL-----OQPVSDGQGDQDSA- 2591
2966 EDDRGAYLCAYVAVGEATPAQLRAYMENELPNYVWPAFFIOLEKMLTPNDKIDRKAL 3025
2592 -----HMAPRTEETAILCDEFAKVLGF-QVGITDNFDFLGGHSLMATKLAURI 2638
3026 PRKQNEENRTEQYAAQPOTELEQLAGIWDVLIQVGTQDNFFELGGDSIKAIQVSTRL 3085
2639 GHRLDITVSVKQVDFHPVLFOALALDNLVQSKTNEIVGCREMAEYSPFOLLTEDEEF 2698
3086 --NASGWTAMKELFOYPTTEEAALRV-----IPNSRE----- 3116
2699 MASEIKPQLEQIIOQIVPSTOMQAKFLFDHTTAR---PRPFVPFYID-FPSTSEPDAA 2754

| | | | |
|------|----|---|------|
| 4827 | Db | AEALDASGTVRTALTARPRERIPLSVYAQORLWFLHOLEGPSATYNVLTLRUGCALDVD | 4988 |
| 1147 | Qy | ALRRALAALBQRHETLTRTFEDQGVQIV--HEKLSEMKVIIDLCGSDLDPEFVNQE | 1204 |
| | | ::: :: : : : : : | |
| 4887 | Db | ALRAAISDVVARHESLRVTFTEDERGAYOIVLPVEAASTPFTVDVAEEIG--DRLEDA | 4944 |
| 1205 | Qy | QTTPFNLSBAGWRATLLRGEDDHILTIWMHHIISDGWSIDVRLRDOLYLSAALKDSK | 1264 |
| | | ::: : : : : : : : : : : | |
| 4945 | Db | VGCFCDLAOPARTSLFRYSEREHWLLLIIHHIASDAWSRAPLAODLTAAAYAARV-SE | 5003 |
| 1265 | Qy | DPLSALTPLPIXYDSFAKWQDOFIQE-----KOLNWKXKOLD-SSPAKIPTDFAR | 1316 |
| | | ::: : : : : : : : : : : | |
| 5004 | Db | APMWA--PLUQVADTALWOEILGDTDADSIAQLAYWKOOGLAGLPOLQDLPDRPR | 5061 |
| 1317 | Qy | PALLSGDAGCVHTIDGELYOSRAFACNEHNITTSFVVLLAAFRAAHYRLTAVEDAVIGTP | 1376 |
| | | : : : : : : : : : : : : | |
| 5062 | Db | PAVAGYSGDRVPTVPTELTRTELARATISAFVMYIOAAVAVLLTRCAGEDIPIGTP | 5121 |
| 1377 | Qy | IANNRPLEDITIGCFNVTCMRINIDHDHDTGCTLINOVKATTATAFENEDIFFERVVSA | 1436 |
| | | : : : : : : : : : : : : : | |
| 5122 | Db | VAGRTDDAADLGIFINTVLVLTDTSGDPFRRLDRVRDRTLAAAYAHODLPFERLVEA | 5181 |
| 1437 | Qy | LQPSRDLSTPLAQLI FAVHSOKDLGRFX---FOGLE-SVPVPSKATRFDMEFHLFQ | 1491 |
| | | : : : : : : : : : : : : : | |
| 5182 | Db | LNP-ARTLSHPHFQVLLTFNNTDHEGALKDISELPCLTVALREVORTSSKFDSLSCFAE | 5240 |
| 1492 | Qy | ETDS-----LKGSVNADELPMKEVENVVVFPEILLRNGLOSSRFPVSILPLTGIVT | 1545 |
| | | : : : : : : : : : : : : : | |
| 5241 | Db | SPTSRRPGIEAALDFSTELLDRRQAQIADRLVKVLEAVTAPDRPIGAVELMDPAER | 5300 |
| 1546 | Qy | LERLDLVNKHVDPRESSLADVFOQSAYPSDLAVDSSCRILTVELDRQSDIILAGWL | 1605 |
| | | : : : : : : : : : : : : | |
| 5301 | Db | ERYLVENWGAPTOLP-GTPLHELISEGARLTPDAVAVCDGTTLTAELDRRANQARHL | 5359 |
| 1606 | Qy | RRRSMAETLVAVFAPRSCETTIVAFQVLKANLAYIPLDVSRPSARVQDILSGLSOPTV | 1665 |
| | | : : : : : : : : : : : : | |
| 5360 | Db | LGEGLAGEDFVALAKSLDAVISMLAVLKTGAAYLPIDPDYPAERTYML----- | 5410 |
| 1666 | Qy | LIGHTAPPDIEVTNVEFVRIRDALNSADGNAGEVTEHDSK-P-SATSLAYLVLYTSGSTG | 1724 |
| | | : : : : : : : : : : : : : | |
| 5411 | Db | ----DDAQPALTLT---APIPPASYDSRPTS-EITDVERSPWARHAAYMIYTSGSTG | 5461 |
| 1725 | Qy | RPKGMIEHRVIRTVTSGCIPNPSETMAHM-ATIAFDGASYEIIYSALLFGRTLVCVD | 1783 |
| | | : : : : : : : : : : : : | |
| 5462 | Db | RPKGVVYIEHA-LATYLHRARNITYTAMGTVTVLHRSPLAFLTITALVTLTAGTIV--- | 5516 |
| 1784 | Qy | YMTLDBARKKVFFREHYNAAHSVTTSSODVPL--RVPRRL--SRTLMEFFLVVTDSTA | 1839 |
| | | : : : : : : : : : : : : : | |
| 5517 | Db | HLSFSLEAEVQ-----PSLIKATPSHLPLLTLTPETASPSHLLI-----G | 5557 |
| 1840 | Qy | PDALDAQGL-----YOGVOCYNGYPTENGVMNSTIYPIDSTESFINGVPIGRALNNSGA | 1893 |
| | | : : : : : : : : : : : : | |
| 5558 | Db | GEALHTDHLATWRTQHPGAQILINAYGPTESTVNITDHHV-SEDTPGVPVPIGRPANTQV | 5616 |
| 1894 | Qy | XVVDPEQOLVGIVMGELVVTGDLARGYSDK-ALDENRFVHITVNDQT-----VKAYRTG | 1948 |
| | | : : : : : : : : : : : : | |
| 5617 | Db | YVLDSALRPVAPGVTGELYLAGELARGYLGRLPALTAEFP---TANPHSSTPGARMYRTG | 5673 |
| 1949 | Qy | DRVRYIGDGLIEFFGRMBDTQFKIRGNRIESAEBIEAALLRDSVRDAAVVVLOONQDAPE | 2008 |
| | | : : : : : : : : : : : : | |
| 5674 | Db | DLAHWN-HDGHLTIDGRADHQIKLRGHRIEPGEIETTLTAQTGITQTTVOLRE----- | 5725 |
| 2009 | Qy | ILGFVVADHDSENDKGQSANQVEGWQHIFESGMSDIGEIDPSTIGSFCKWTSMYDGS | 2068 |
| | | : : : : : : : : : : : : | |
| 5726 | Db | -----DTPG----- | 5729 |
| 2069 | Qy | QIDFDEMHEWLGTTTRLTHDNRSGLNVLEIGTSGMILFNLDLSRLESYVGLPEPSRAAAF | 2128 |
| | | : : : : : : : : : : : : | |
| 5730 | Db | -----DQRLVAYL-----V | 5738 |
| 2129 | Qy | VNKATESIPSACKAKVQGTATDIGQVDLHPDLVVLNSVIOYFPSSEYLAETDLTH | 2188 |
| | | : : : : : : : : : : : : | |
| 5739 | Db | VNDSTE----- | 5744 |

RESULT 15
C69681
peptide s

N:Alternate names: peptide synthetase 3

C:Species: *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence revision 05-Dec-1997 #text_change 03-Nov-2000

C:Accession: C69681; I40458; S49135

R:Kunst, F.; Ogasawara, N.; Woszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho

A.; Enrich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois

A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Odawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serch

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. *Bacillus subtilis*.

A:Title: The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: C69681

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2555 <KUN>

A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CABL3715.1; PID:g2634215

A:Experimental source: strain 168

R:Tognoni, A.; Franchi, E.; Magistrelli, C.; Colombo, E.; Cosmina, P.; Grandi, G.

Microbiology 141, 645-648, 1995

A:Title: A putative new peptide synthase operon in *Bacillus subtilis*: partial characteri

A:Reference number: I40454; MUID:95227362; PMID:7711903

A:Accession: I40458

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-859 <RES>

A:Cross-references: EMBL:Z34883; NID:g509465; PID:g509470

C:Genetics:

A:Gene: *ppsC*; *pps3*

C:Superfamily: peptide synthetase ppsD; acetate-CoA ligase homology; acyl carrier protei

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein.

F:510-953/Domain: acetate-CoA ligase homology <ACLI>

F:970-1038/Domain: acyl carrier protein homology <ACP>

F:1551-1950/Domain: acetate-CoA ligase homology <ACL>

F:2006-2073/Domain: acyl carrier protein homology <ACP2>

F:1002-2038/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 11.2%; Score 1807.5; DB 2; Length 2555;

Best Local Similarity 24.7%; Pred. No. 9e-97;

Matches 681; Conservative 510; Mismatches 1114; Indels 457; Gaps 101;

QY 46 IEAKPCTPFQDMIDCNALDKQSA--IGHAVYDVPDIDISRFAALWKEIVNQTPALRA 103

Db 7 IQDIYPLSPFQEGMLPHSLYDEQSRAYFEQASFTIHGQDLERFKSMDAIVEDRYDIFRT 66

QY 104 FAFSTDSGKTSQVILKDSFV-----FSWCMWSSSSPDEVRDEAAAAAGPRCNRFLV 157

Db 67 AFTYKNVAKPRQVVLKQRCPTIHIEDISHLNRKDEKHECTAEKFEQ-----DKSKGF 119

QY 158 LEDMQTKKQQL-----VMTFSHALVDVTFQQRVLGRVFAAYKHEKDHTRPTEPSS 208

Db 120 QTDVLMRISLTKWAPDHYVCINSHHLLMDGWCGLGIVIKDFLHIYQALCKGQDLPLPPVQ 179

QY 209 DATFDTDSQSVSVVSMCEDNAVSAATHFWQTHLNDLNASV--FPHLSOHLMPVNPPTTAHR 267

Db 180 PYG-----TYIKLWLMQDREAEAYWKRLQHEKSTPLPKRTD--QIPNGTL--QQ 227

QY 268 IFPPLSQK-----ALNSAICRTAL-----SILLRSTHSDEALFGAVTESLPFDKH 315

Db 228 ITFAIPEKETAELOKTAASAGATLNTVFQALWGLMLOKVNRRSSDAVFGSVI--SGRPE-- 284

QY 316 YLAD-----GTYQTVAPLRVHCOENLRASDV-----MDAISY----- 348

Db 285 -LKDVENMVGFLNTIPTIRAQSDS--LSFSDLVRRMQKMDNEAEAYSFFLYDITQAQSA 342

QY 349 DDRLGHLAPF-----GLRDIRNTGNGSAAACDFQTVLLVTDGSHVNGINGFLQOITE 401

Db 343 QELIDHIIIFENTPTQOEIEELNAQGSDFSVKDFE-----MEEVTN 384

QY 402 SSHFMPCN-----NRALLLHCQMESSGALLYVDHNVDSLOTTRLLQOFGHLIKCQS 456

Db 385 YS-----CSVKVIPGRLYVRIHQFQSA-----YQPSMMSEIKDYLL-----HMVSDVIS 429

QY 457 PDLSSMAEVLNMTEDYRAEIESWNSQPLEV--QDTLIHHEMLKAVSHSPKTKTAIAOWG 514

Db 430 DPSL-PVSKMTLLDEDKTRKIVSONNRTYSVSFEPTLHGLFERQAAVTPERLAIRPSGG 488

QY 515 DWTYSELDNVSSRLAVHIIKSLGRQAQAIIPVYFEKSKWVIASMLAVLKSGNAFTLIDPN 574

Db 489 SLTYAELDMYASRLAAHLAARGV-TNESIVGLSERSPDMLIIVLAVLAKGAGAYLPDPA 547

QY 575 DPAPTAQVVTQATRAVATSKLHRETQKLVGRVGVVVDDELLOSVSASDDFSLTKSQD 634

Db 548 YPKERUSYMLUKDSGASULLLTP--GCSAPNFSGETLEVDWMTSCECEVKRHSVSAVSQDGS 605

QY 635 LAYVFTSGTGDPKGMIEHRAFSFSCALFKFGASLGINSIDTRALQFQTHAFGACLLFMT 694

Db 606 LAIVIYTSGTGPKGVAVEHQAVSFLTGCMQHOFRLSEDDIVMVKTSFSDASVWQLFW 665

QY 695 TLINGCVCI--PSDDDRMNSIPSFIRNVNVMNMATPSYMGTFSP--DVPGLA 745

Db 666 WALSGASAVLLPPGWEKDSALIVQAIHQENVVTAHFIPAMLNSFLDQAEIERLSDRSLK 725

QY 746 TLVLVGEOMSSSVNATWA---PKLQLLNGYGOSESSICFASNMSTEPNN-----MGRA 796

Db 726 RVFAGGEPLAPRTAARFASVLPQVSLHNGYGPTEAVTDAAFVYVLDPERDRDRIRIPIGKP 785

QY 797 V-GAHSWVIDPNDINRLVPTGAVGELVIESPGIARDYIVPPPEKSPFFTDIPSWYPANT 855

Db 786 VPCARLYVLDPH--LAVQPSVAGELYIAGAVGYLNRPAUTEERFLED-----P 835

QY 856 FPDGAKLYRTGDLIARVASDGSIVCLGRIDSQVIRGORVELGAETIHLROOMPDDLTIIV 915

Db 836 FYLGEMYITGQDVARWLPDGNVFLGTDDQVQIRGYRIEPEIEAALR-----SIEGVR 890

QY 916 EA--TKRSOSANSTSLIAFLIGSSYFCGNRPSDAHILDHDKATKINIKLEQVLPKHSIPSF 973

Db 891 EAAVTVRTDS-GEPELCAYVEG-----LORNEVRA---QLQRLPGYWPAY 933

QY 974 YICMLELPRTATGKIDRRRLRMKGDILDKQTQCAIVQQAPAPIPVFAADTAALKHSIWVQ 1033

Db 934 MIEMEQWPTPSCKLDRNALPAPG-GAADAET-----YTAPRNV-----TEMKLSQMED 982

QY 1034 SLGIDPATVNWGATFFELGNSITATKMN--MARSGMDLKVSNIYQHTPLAGISAVVK-- 1091

Db 983 VLKNGP--VGIDHNFDRGGHSLKATALVSRIAKEFDVQVPLKDVFAHTVEGLATVIRE 1040

QY 1092 GDPLSYTLIPKSTHEGPVEQSYSGRLWFLDQLDVGSLWLYLIPYAVMRGPNVNDALRRA 1151

Db 1041 GTDSPYEATKPAEKQETPVSSAQKRIYVIOQLEDGGTGYNMFAVLEGLKLNLRMDRA 1100

QY 1152 LAALQORHETRLTTFE--DQDGVGVQIVHEKLSSEMKVIDLCGSDLDPEVLNQDQT--- 1207

Db 1101 FKELIKHESLRTAFEDQAGDGPVRIHDEV-----PFTL---QTTVLG 1141

QY 1208 -----PFLNLSSEAGWRATLLRLGEDDHILTIYVHHIISDGSIDVLRDLNQ 1254

Db 1142 ARTEEEAAAFAIKPFDLSQAPLFRQAIVKVSDEHLLLDVDMHHIISDGSVSNILIREFGE 1201

QY 1255 LYSAAALKDSKDLPSALTPLPIQYSDFAKWK-----DOFIEQKQLNWKMKOLKSSPA-- 1308

Db 1202 LYNN-----RKUPALRIQYKDVAVMOEGFKTGDAY---KTQGGAYWLKQEGEGLPV 1249

QY 1309 KIPDTFARPALLSGDAGCVHVTIDGELYQSIRAFPCNEHNTTSFVVLIAAFRAAHRILTAV 1368

Db 1250 DLPADHARPPWRSPFAGDKVSTLDQEVTSGLYKLARENGSTLYNVLLAAAYTAFSLRSQG 1309

QY 1369 EDVIGTPIANRRPELEDIIGCFVNTQCMRINIDHHDFTGTGLINQVKAATTTAAFENEDI 1428

Db 1310 EDIIVGSPDIAGRPHKDEPIILGMEVNTIALTRPEGKPFVQYLOEVRETAMEAFERQDY 1369
QY 1429 PFERVVSALOPGSRDLSTPLAQILFAVHSOKDLGRKFQGLSVPPVPSKAY--TRDME 1486
Db 1370 PFEELVOKLEL--TRDSNRNPLFDVMEVLQNM--DOESLELDELCLKPRAANNGHOTSKEDLT 1427
QY 1487 PHLFOETDS--LKGSVNFADELFKMETENVVVRVFEILNRGLQSSRTPVSLPTDGIIVT 1545
Db 1428 LYAQEOGRLLTFQMEFSTDLKKYKTKWLOYENNNLLSIKONKAALGTINILNEDEA 1487
QY 1546 LEKLDLVNKHVDYPRSSADLVFOTOVSAYPDSLAVVDSCLTYTIEDRQSDILAGWL 1605
Db 1488 HYLHELNRKIDIPRNETISRLFEMQAEQTPNAVAIVSDQVFTYEDLNSWANQIASVL 1547
QY 1606 RRRSMPAETLVAVPAPRSCETIVAFFGVGLKANLAYPLDVRPSARQDILSGLSGPTIV 1665
Db 1548 QIKGVGPDVVALLTGTPELJAGMLGILKAGGAYLPIDSNLPVERTAYMLS--DSRAAL 1605
QY 1666 LIGHDTAPPDIEVNVFEVRIRDALNDSNADGFEVIEHDSKPSATSLAVVLYTSGSTGR 1725
Db 1606 LLOSEKTEKRLUGIECEQIIIEDTQKQGEAKNVE--SSAGPH--SLAYIYTSGSTGK 1659
QY 1726 PKGVMIHRVIRIVTSCIPNYPSETRMAHMATIAFDGASYEYISALLFCRTLVCVDYM 1785
Db 1660 PKGVMIORSVIRLVKNSYITFTPEDKLLMTSSIGFDVGSFEIFGLLNGAALHLSDDQ 1719
QY 1786 TTLARALKDVFFREHVNAASHVTSQDQVPLVRPRLSLTMEFFLVVTDSTAPDALD- 1844
Db 1720 TFLDSHQLKR--YIEHQGITIMLTSSLEFNLHTEQNEOTFSQLKHLIIGCEALSPSHVR 1777
QY 1845 AGLYQGVQCYNGVGTENGVMSTIYIDSTESFINGVPIGRALNLSGAYVVDPEQOLVG 1904
Db 1778 IRNVCPEVSNWNGVGTENTFTFSLHIQKTYEL--SIPICRPVGNSTAFILNQGVLQP 1835
QY 1905 IGVMGELVWTDGLARGYSKA--LDENRFV--HITVNDQTVKAYRTGDRVRYRIGDGLIE 1961
Db 1836 VGAVGELCVGDGVARGYLGRPLDTEKFPHPAPGD--RLYRTGDLARW--LSDGTIE 1891
QY 1962 PFGMDTQFKIRGNRIESAIEAALLRDSVVRDAVVLQONEDQAEILGVVADHDHSE 2021
Db 1892 YVGRIDQVQVRYGVELGETALRQIDGVEAAVLARTAQTKGSKELFGYISV----- 1945
QY 2022 NDKQOSANOVE-----GWQDHESGMYSDIGEIDPSTIGSDFKGWTSMYDGS 2068
Db 1946 -KAGTNAEQVRLARSLLPNMIPAYIEMETPLTNSGNLKRKALPEPDVASKQTYIPP 2004
QY 2069 QIDFDEN-----HEWLGETTRTLHNRSLGNVLEIGTSGMILFNLDLSRESY--VGLEPS 2122
Db 2005 RNELEEQLALIWQEVLGIGRIED-----SFFELG--GDSIKALQVSARLGRYGLSLOVS 2058
QY 2123 -----RSAAAFVNKATESIPSILAGKAKVQGTATDIGO-----VDDLHPDLVVLNS 2168
Db 2059 DLFRHPRIKIDLPFIRKSERIIEQ--GPIQGDV--PWTVPQWPFSSQDIEERHH-----FNQ 2111
QY 2169 VIQVF-----PSSEVLAEIADTLIHLPNVQRIFFGDVR-----SQATNEHLAA 2212
Db 2112 SVMLFHSGRLSENALRPAKLAHHHDAL-----RMVYRNDRRWTQINOIHESOLYS 2165
QY 2213 RAIHTLGKNAT--KDDVROKMAELE-----DMEELLVEPAFTSLKDRFPGLVEH----- 2261
Db 2166 LRISDLQSSEGWETKIQEVAADLQOSINLOEGPLLHAALFKTLTGDYLEFLAIHHLVVDG 2225
QY 2262 --VEILPKMNEAVNELSAYRYAAVHVVRVSGDELVLPEKDDWIDFQANQLNOKSLGDL 2319
Db 2226 VSWRILLELSA-----GYQOAAA-----GOTIQLPKPTDSYQY--ARRIQEYAGSSK 2272
QY 2320 LKSSDAIMAV-----SKIPFEITAFERQVVASLNSNIDEWQLSTIR--SSAEGDSSLSVP 2373
Db 2273 LIREAYWRSVEQQAELPYEIPH-----HVNIDFSKRDSLSFSLSEADTAVLLQ 2323
QY 2374 DIFRIAGAGFRVEVSSAR-----QWSONGALDAVFHHCSSQGR----- 2412

Db 2324 NVNHAYGTDQDIIILLTAASLAICEWGTGSGKLRIAME---GHGREHILPELDISRTVGWFT 2380
QY 2413 ----TLVNFPDTHHLRGSDLLTNR--PLQRLQNRRIAEVRRERLSLLPSYMIPSN--IVVL 2466
Db 2381 SMYPALISFENHRDELGTISVKTVDTLGRIPNKGVGY---GMLKYLTHPENKSTTF 2433
QY 2467 DKMP--LNANGVKDRKELSSRAKVVPKQOATAAPLPTFPPISEVEVILCEBATEVFGMKVDIT 2525
Db 2434 SKTPEISFNILGQFNDIERQDTERPSSILGSGKDITHTWKREQII-----EWSAMAADKK 2487
QY 2526 DHFENLG-----GHSLLATKLISRIDQRL-----KVRITVKDVFDBHPVFA-DIA 2568
Db 2488 LH-FNLSYPPARFHRNTMEQLINRIEHFLDIMKHCAGQQAECTLSDFSQSLTAEDLD 2546
QY 2569 SV 2570
Db 2547 SI 2548

Search completed: May 30, 2003, 12:50:02
Job time : 164 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protéin - protein search, using sw model

Run on: May 30, 2003, 12:39:11 ; Search time 161 Seconds
(without alignments)
4004.481 Million cell updates/sec

Title: US-09-482-788-2
Perfect score: 16128
Sequence: 1 MEYLTAVDGRQLPPTPASF.....RVEHLLLEVSKTPEGLNSSL 3129

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPREMBL_21:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriaph:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 9874 | 61.2 | 3131 | 3 Q00869 | Q00869 fusarium eq |
| 2 | 5000 | 31.0 | 15281 | 3 Q09164 | Q09164 tolypocladi |
| 3 | 2871 | 17.8 | 983 | 3 Q00868 | Q00868 gibberella |
| 4 | 2496 | 15.5 | 1051 | 3 P97961 | P97961 cylindrotri |
| 5 | 2391 | 14.8 | 4247 | 2 Q918H4 | Q918H4 streptomyc |
| 6 | 2275.5 | 14.1 | 4848 | 2 Q07944 | Q07944 streptomyc |
| 7 | 2146.5 | 13.3 | 2841 | 2 Q9FB33 | Q9FB33 streptomyc |
| 8 | 2108 | 13.1 | 5060 | 2 Q9K5M1 | Q9K5M1 anabaena sp |
| 9 | 2018.5 | 12.5 | 5157 | 3 Q01135 | Q01135 metarhizium |
| 10 | 2002 | 12.4 | 4360 | 3 Q9UVN5 | Q9UVN5 alternaria |
| 11 | 1984 | 12.3 | 1997 | 2 Q05647 | Q05647 streptomyc |
| 12 | 1978.5 | 12.3 | 3317 | 2 Q9RAH2 | Q9RAH2 nostoc sp. |
| 13 | 1974 | 12.2 | 4379 | 2 Q9RAH4 | Q9RAH4 nostoc sp. |
| 14 | 1933 | 12.0 | 4450 | 2 Q44928 | Q44928 bacillus br |
| 15 | 1906.5 | 11.8 | 3670 | 16 Q924X5 | Q924X5 streptomyc |
| 16 | 1896.5 | 11.8 | 2588 | 16 Q8YTR5 | Q8YTR5 anabaena sp |

| | | | | | |
|----|--------|------|------|-----------|--------------------|
| 17 | 1892.5 | 11.7 | 6889 | 16 Q8XS40 | Q8XS40 ralstonia s |
| 18 | 1880.5 | 11.7 | 2450 | 2 Q9RAH1 | Q9RAH1 nostoc sp. |
| 19 | 1879 | 11.7 | 5953 | 16 Q8XS39 | Q8XS39 ralstonia s |
| 20 | 1876.5 | 11.6 | 9376 | 2 Q85168 | Q85168 pseudomonas |
| 21 | 1867.5 | 11.6 | 5149 | 16 Q91179 | Q91179 pseudomonas |
| 22 | 1866.5 | 11.6 | 3588 | 2 Q66070 | Q66070 bacillus li |
| 23 | 1860 | 11.5 | 3310 | 2 Q9AMR5 | Q9AMR5 bradyrhizob |
| 24 | 1831.5 | 11.4 | 3583 | 2 Q45675 | Q45675 bacillus su |
| 25 | 1828 | 11.3 | 2258 | 2 Q9K5M2 | Q9K5M2 anabaena sp |
| 26 | 1811 | 11.2 | 3589 | 2 Q69246 | Q69246 bacillus li |
| 27 | 1810 | 11.2 | 7463 | 16 Q924X6 | Q924X6 streptomyc |
| 28 | 1803 | 11.2 | 2554 | 2 Q30981 | Q30981 bacillus su |
| 29 | 1795 | 11.1 | 3316 | 2 Q8VQF8 | Q8VQF8 xenorhabdus |
| 30 | 1781.5 | 11.0 | 2617 | 16 Q8YTR9 | Q8YTR9 anabaena sp |
| 31 | 1755 | 10.9 | 2571 | 2 Q87704 | Q87704 bacillus su |
| 32 | 1729.5 | 10.7 | 5369 | 2 Q9R9J0 | Q9R9J0 bacillus su |
| 33 | 1725 | 10.7 | 3603 | 16 P94459 | P94459 bacillus su |
| 34 | 1715 | 10.6 | 4898 | 2 Q93N87 | Q93N87 streptomyc |
| 35 | 1713 | 10.6 | 2448 | 16 Q91182 | Q91182 pseudomonas |
| 36 | 1713 | 10.6 | 5362 | 2 Q93155 | Q93155 bacillus su |
| 37 | 1708.5 | 10.6 | 2448 | 2 Q51338 | Q51338 pseudomonas |
| 38 | 1686.5 | 10.5 | 2246 | 2 Q9AK56 | Q9AK56 pseudomonas |
| 39 | 1677.5 | 10.4 | 3582 | 2 Q66059 | Q66059 bacillus li |
| 40 | 1677 | 10.4 | 4342 | 16 Q91157 | Q91157 pseudomonas |
| 41 | 1672 | 10.4 | 2611 | 2 Q68487 | Q68487 streptomyc |
| 42 | 1649.5 | 10.2 | 2378 | 2 Q9R912 | Q9R912 bacillus su |
| 43 | 1634.5 | 10.1 | 499 | 3 Q96UG8 | Q96UG8 fusarium pa |
| 44 | 1629.5 | 10.1 | 3232 | 3 Q94205 | Q94205 claviceps p |
| 45 | 1625 | 10.1 | 3583 | 2 Q45295 | Q45295 bacillus li |

ALIGNMENTS

RESULT 1

| | | | | | |
|--------|---|---|--------------|------|----------|
| Q00869 | ID | Q00869 | PRELIMINARY; | PRT: | 3131 AA. |
| AC | Q00869; | | | | |
| DT | 01-NOV-1996 | (TREMBLrel. 01, Created) | | | |
| DT | 01-OCT-2000 | (TREMBLrel. 15, Last sequence update) | | | |
| DT | 01-MAR-2002 | (TREMBLrel. 20, Last annotation update) | | | |
| DE | Enniatin synthetase. | | | | |
| GN | ESYN1. | | | | |
| OS | Fusarium equiseti. | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; | | | | |
| OC | Hypocreales; mitosporic Hypocreales; Fusarium. | | | | |
| OX | NCBI_TaxID=61235; | | | | |
| [1] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=LAMBOTTE ET FAUTREY; | | | | |
| RX | MEDLINE=93247491; PubMed=8483420; | | | | |
| RA | Haese A., Schubert M., Herrmann M., Zocher R.; | | | | |
| RT | "Molecular characterization of the euaitiu synthetase gene encoding a | | | | |
| RT | multifunctional enzymecatalysing n-methyl depsiptide forluation in | | | | |
| RL | Fusarium scirpi.; | | | | |
| RL | Mol. Microbiol. 7:905-914(1993). | | | | |
| [2] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=LAMBOTTE ET FAUTREY; | | | | |
| RA | Zocher R.; | | | | |
| RN | Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases. | | | | |
| [3] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=LAMBOTTE ET FAUTREY; | | | | |
| RA | Zocher R.; | | | | |
| RL | Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases. | | | | |
| DR | EMBL; Z18755; CAA79245.2; - | | | | |
| DR | HSSP; P14687; IAMU. | | | | |
| DR | InterPro; IPR000873; AMP-bind. | | | | |
| DR | InterPro; IPR001242; Condensatn. | | | | |
| DR | InterPro; IPR003880; Prantne_attach. | | | | |
| DR | Pfam; PF00501; AMP-binding; 2. | | | | |
| DR | Pfam; PF00668; Condensation; 2. | | | | |

DR pfam; PF00550; pp-binding; 3.
DR PROSITE; PS00075; ACP_DOMAIN; 6.
DR PROSITE; PS00455; AMP_BINDING; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHE
KW phosphantetheine.
SQ SEQUENCE 3131 AA; 346494 MW;

```
QY 2024 KQSANQVEGQDHFESGMYSDIG-EIDPSTIGSDFKGWTSYDQSGIDFDEMHFWLGET 2082
Db 2013 EEETGNQVEGQDHFESGMYSDISTAVDQSAIGNDFKGWTSYDQSGIDKQGEWLLDDA 2072
QY 2083 TRTHLNRSLGNVLEIGTSGMILFNLDLSRLSVGLPSPRSAAAFVNKATESPISLAGK 2142
Db 2073 IHTLHNGQIPRVLEIGTSGMILFNLPNGLSVGLDPSKSAVEFVNRAVRESSPKFAGK 2132
QY 2143 AKVOVGATDIGOVDLHPDLAVLNSVIOYPPSSPEYLAELADTLHLPNVORIFFGDVRS 2202
Db 2133 AKVHGMATDVKLGEVHPDVLVFNSSVQYFPEYLAELVIGLGLAIAPSVKRIEGLDLS 2192
QY 2203 QATNEHFLAARAIHTLG--KNATKDDVYRQKMAEELDEEELVPEPAFTSLKDRFPGLVE 2260
Db 2193 YATNGHFLAARAIHTLGTNNATKDRVRQKIOELEDREELVPEPAFTSLKERRPDVVK 2252
QY 2261 HVEILPKMNEAVNELSAYRYAAVHVVRGSLDELVLVPEKDDWIDFQANQLNQSLGLDL 2320
Db 2253 HVEILPKMKNATNELSAYRYAVVHLRDET-DEPVYHIEKDSWVDFAKQMDKTALLDHL 2311
QY 2321 K-SSDAATMAYSKIPFETAFERQVVASLNSNIDE-----WOLSTIRSAEGDSS 2369
Db 2312 RLSKDMASVAVSNITYAHTAFERLIVESLDESDKDTKGLDGAW-LSAVRSAAENRAS 2370
QY 2370 LSVPIDFRIAGEAGFREVSSARQWSONGALDAVFHHC--CSQGRTLVNFPTDHHLRGSD 2427
Db 2371 LTVPDILEIAKEAGFRVESAARQWSONGALDAVFHFPSPSTDRTLQIOPPTDNLRSLL 2430
QY 2428 LITNRPQRLQNRRIATEVRELRSLPSYMPNSIVLDKMLNANGKVBKELSRRAK 2487
Db 2431 TLANRPQKLRRAALQVREKLOTLPYSYMPNIVLDVMTPLNTNGKIDKELTRRAR 2490
QY 2488 VVPKQOTAAPLTPPISREVEVLCEATEVFGMKVDITDHFENLGGHSLATKLISRIDQ 2547
Db 2491 TLPKQOTAAPVDPPISDIETILCEATEVFGMKVEISDHFELGGHSLATKLISRLOH 2550
QY 2548 RLKVRITVKDVFDPHVPFADLASVIRQGLGLOQPSVSDQG--QDRSAHMAPRTETEALCDE 2606
Db 2551 RLHVRVTVKDVDFSPFADLAVIIRQGLAMQNPVAEGDKQWGSRAVPRTVEKMLCEE 2610
QY 2607 FAKVLGFOVGTIDNFFDLGGHSLMATKLAVRIGHRLDITVSVKQVDFH-----PVLQ 2659
Db 2611 FAAGLVGVGTIDNFFDLGGHSLMATKLAVRIGRL-----IRHSHQGLRLPCAPQ 2662
QY 2660 LAIALDNLVQSKTNEIVGREGMAEYSPQLLFTDPEEFMAEIKPOLE-LQEIITQIYVP 2718
Db 2663 LAKLESSHKSYESGDDIQADYTAFQLDLEDQDFVQSQIRPQLDSCYGIQIYVP 2722
QY 2719 STQMKAFLEDDHTTARPPFPVFPYIDFPSTSEPDAAGLIKACESLVNHLDIRTVFAPAS 2778
Db 2723 STQMKAFLEDDHTTARPPFPVFPYIDFPSTSEPDAAGLIKACESLVNHLDIRTVFAPAS 2780
QY 2779 GELVQVVLSCLDLPLOVLEIETEDNTATNEFLDEFAKEPVRGLHPLIRFTIILKQTKSMRV 2838
Db 2781 GDLVQVVEHLNPLIETIETKNNWTATGDYLDVHGKDPVRLGHPCIQIFAILKTASVRY 2840
QY 2839 IMRISHALYDGLSLHVVYRKLHMLYNGSLPPOHFSRYMOYVADAGRESHGFWRDVION 2898
Db 2841 LLRSHALYDGLSLHVVYRKLHMLYNGSLPPOHFSRYMOYVADAGRESHGFWRDVION 2900
QY 2899 TPTMTILSD--DTVVVDGNDATKALHLSKIVNIPSOVLGSSNIITQATVFNAACALVLSR 2956
Db 2901 APMTVLHTDNTNGMSQEPASKAVHLSEVNVNPAQAIRNSTN--TQATVFNTACALVLAK 2958
QY 2957 ESDSKDVYFGRIVSRQGLPVEYQDIVGFCPTNAPVRAHIESSDYNQLLHDIQDYLSSL 3016
Db 2959 ESGQDVYFGRIVSRQGLPVEYQDIVGFCPTNAPVRAHIESSDYNQLLHDIQDYLSSL 3018
QY 3017 PHETIGFSDLRNCTDWEATITNFSCCITTYHNFYHPESQEQOQVEMGVLTILKFNVIEMD 3076
Db 3019 PFESLGFBEIRNCTDWEATITNFSCCITTYHNFYHPESQEQOQVEMGVLTILKFNVIEMD 3078
```

```
QY 3077 EPLYDLATAGEVPEPDGAGLVKVTYIAKTQLFGRKRRVHLEEVSKTFEGNSSL 3129
Db 3079 EPLYDLATAGEVADGVNLKVTYVAKARLYNEARIRHVEEVCKTFNGINEAL 3131
RESULT 2
Q09164 PRELIMINARY; PRT; 15281 AA.
AC Q09164;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Cyclosporin synthetase (CYSYN) (EC 6.-.-.-).
GN SIMA.
OS Tolypocladium inflatum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae;
OC Tolypocladium.
OX NCBI_TaxID=29910;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 34921;
RX MEDLINE=95094306; PubMed=8001164;
RA Weber G., Schoerndorfer K., Schneider-Scherzer E., Leitner E.;
RT "The peptide synthetase catalyzing cyclosporine production in
RT Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading
RL frame.";
RL Curr. Genet. 26:120-125(1994).
CC -!- FUNCTION: THE CONSTITUENT AMINO ACIDS OF CYCLOSPORINS ARE
CC ACTIVATED AS AMONACYL-ADENYLATES WITH PEPTIDE BONDS FORMED
CC THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER INTERMEDIATES.
CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC -!- COFACTOR: CONTAINS 11 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF CYCLOSPORINS.
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
DR EMBL; 228383; CAA82227.1; -.
DR HSSP; P14687; IAMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR000267; Asp/Glutamase.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Pantne_attach.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF00501; AMP-binding; 11.
DR Pfam; PF00668; Condensation; 13.
DR Pfam; PF00550; PP-binding; 11.
DR PRINTS; PR00154; AMPBINDING.
DR PRINTS; PR00139; ASGNLASE.
DR PROSITE; PS50075; ACP_DOMAIN; 11.
DR PROSITE; PS00455; AMP_BINDING; 10.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_9.
KW Ligase; Antibiotic biosynthesis; Phosphopantetheine; Repeat;
KW Multifunctional enzyme.
FT DOMAIN 15179 15219 13 X 3 AA APPROXIMATE REPEATS.
FT REPEAT 3 1086 DOMAIN 1.
FT REPEAT 1087 2585 DOMAIN 2.
FT REPEAT 2586 4072 DOMAIN 3.
FT REPEAT 4073 5564 DOMAIN 4.
FT REPEAT 5565 7061 DOMAIN 5.
FT REPEAT 7062 8121 DOMAIN 6.
FT REPEAT 8122 9616 DOMAIN 7.
FT REPEAT 9617 11113 DOMAIN 8.
FT REPEAT 11114 12185 DOMAIN 9.
FT REPEAT 12186 13681 DOMAIN 10.
FT REPEAT 13682 14767 DOMAIN 11 (ALA-ACTIVATING).
FT BINDING 1060 1060 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 2558 2558 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 4045 4045 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 5537 5537 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 7034 7034 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 8094 8094 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 9589 9589 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 11086 11086 PHOSPHOPANTETHEINE (POTENTIAL).
```

| | | | | |
|---|----------|---|-------|-----------------------------------|
| FT | BINDING | 12158 | 12158 | PHOSPHOPANTHETHEINE (POTENTIAL) . |
| FT | BINDING | 13654 | 13654 | PHOSPHOPANTHETHEINE (POTENTIAL) . |
| FT | BINDING | 14729 | 14729 | PHOSPHOPANTHETHEINE (POTENTIAL) . |
| SQ | SEQUENCE | 15281 AA; 1689051 MW; E26DA7AA35324C05 CRC64; | | |
| Query Match | | | | |
| Best local Similarity 32.4%; Pred. No. 1.2e-299; | | | | |
| Matches 1233; Conservative 440; Mismatches 867; Indels 1264; Gaps 66; | | | | |
| QY | 482 | SQPLEVQDTLIHHMKLVASHSPKTAJQAWGDWTSYSELDNVSSRLAVHIKSLGLRAQQ | 541 | |
| Db | 11581 | NQDYPFADASVIDVFREQVASIPKSTAVIDAASSQLTYTELDERSSOLATWLRR-QVTYPE | 11639 | |
| QY | 542 | AIIPVYFEKSXWVIASMLAKSGNAFTLIDNPDPARTQAQTOT--RATVALTS---- | 595 | |
| Db | 11640 | ELGVLPAPSCETIIFALGIIKANLAYLPDVNAPAGRIETILSSLPGNRLLIGSDTQA | 11699 | |
| QY | 596 | KLHRETQVKLCRCVVVDDELLOSYS-ASDFESSLKQSODLAYVIFTSGSGDGPKGIMI | 653 | |
| Db | 11700 | VKLHANSV-----RFTRIISDALVESPPTELSRTPTAOSLAYVMFTSGSYGVPKGVMV | 11754 | |
| QY | 654 | EHRAFSSCALFKFCAGLSINSDTRALO-----FGTHAFGACGLEIMTTLNGG--CV | 702 | |
| Db | 11755 | EHRGIT-----RLVKNSNVVAKOPAAAAIAHLNIAFDASSWEIYAPLLNGGTVCVI | 11806 | |
| QY | 703 | CIPDDDRMNSPSFINRYNNVMMATPYMG---TFSPEDYPLGLATLVLGEOSSSVN | 759 | |
| Db | 11807 | DYTTID-IKALEAVFKQHIRGAMLPPALLKQCLVSAPTMISSLEILFAAGDRLSQ-D | 11864 | |
| QY | 760 | AIWAPKL---OLLNGYGQSESSICFASNW-STEPNMGRVAG-----AHSWIDPDNDIN | 810 | |
| Db | 11865 | AILARAVSGVYNAYGPTEYVLSTIHNIGENEAFNSGVPAGNAVSNAGFVMDQN--Q | 11922 | |
| QY | 811 | RLVPIGAVGELVIESPGIARDXIIVPPPEKSPFTDIPSMPANTF---PDG--AKLYR | 864 | |
| Db | 11923 | QLVSAGYIGELVVGDLARGY-----TD--SKLRVDRFIYTTLDCGNRVRAYR | 11968 | |
| QY | 865 | TGDLARY-ASDGSIVGLGRIDSQVKIRGORVELGAJETHL-RQOMPDDLTIVVEATKRQ | 922 | |
| Db | 11969 | TGDRVRHPKDGQIEFFGRMDQOIKIRGHRIEPAEVEQALARPADAISDSAVITQLTDBEE | 12028 | |
| QY | 923 | -----SANSISLAFLIGSSYFCGNRPSSDAHILDHDATKA-----INIKLEQVL | 965 | |
| Db | 12029 | PELVAFFSLKNGANNGV-----NGVSDQEIKDGDQHALLMENKIRHNLOALL | 12078 | |
| QY | 966 | PRHSIPFYICMLELPRTATGKIDRRRLRMKDIKDQTOGAIVQOAPAPIPVFADTAA | 1025 | |
| Db | 12079 | PTWIERSRIHVQLPNANGKIDRNELAVR-----AQATPRTSSVSYVAP | 12125 | |
| QY | 1026 | K-LHSIVQSLGIDPATVNWGAT--FFELGGNSITAIRM-VNMARSVGMJDKLVNIOYH | 1080 | |
| Db | 12126 | RNDIETIICKEA-DILSVRVGITDNFFDLGGHSLIATKLAARLSRRLDTRYSVRDFT | 12184 | |
| QY | 1081 | PTLAGISAVVKGDPLSYTLIPKSTHCPCPEQSYSQGRWFLQDLVGSLWYLIPYAVMR | 1140 | |
| Db | 12185 | PVYGQAAAISQOGSTPHEIPAIPALSHSGPVOQSFAGORLWFLDRFNLANAAYIMPFGVLR | 12244 | |
| QY | 1141 | GPVNDALRRALAALBQRHETLRTTPEDODGVCQIVKEKLSSEMKVIDLCSGLDPPREV | 1200 | |
| Db | 12245 | GPLRVDAQTLRALERHELLTTFEEOOGVGMIQVHPSPRMRDICVVDISGANED-LAK | 12303 | |
| QY | 1201 | LNOEQUTPFNLSEAGWRATLRLRGDEDDHILTIVMHIIIISDWSIDVLRNLQLYSAAL | 1260 | |
| Db | 12304 | LKEEQOAPPNLSTEVAWMRVALKAGENHHILSIIVMHIIISDWSVDIFIQELAQFYSVAV | 12363 | |
| QY | 1261 | KDSKPALSALTPLPIQYDFAKWKD--OFIQEKQLNWKWKOKDSSPAKITPTDFARPA | 1318 | |
| Db | 12364 | R-GHDPUSQVKPLPIHYRDFAVNQRODKQAVHESOLOQWIEQLADSTPAELISDNRPE | 12422 | |
| QY | 1319 | LLSGDAGCVHVTIDGELYOSLRAFCEHNTTSFVULLAAFRAAHRYELTAVEDAVTGTPIA | 1378 | |
| Db | 12423 | VLSGEAGTVPIVEDIEYEKLSLFCRNHOVTSEFVULLAARFAHRYELTGAEDATTIGTPIA | 12482 | |


```
Db 164 YVMPARITLIDOMPLNANGKVRKDLARRAQTYSK--ABKLP SARVAPRNEVEVCEEF 221
QY 1021 ADTAALKHSIWQSLGIDPATVNVGATFEELGNSITAIKM-VNMARSVMGLKYSNIYQ 1079
Db 222 SDV-----LGEV---VGVADSFDLGGHSLMATKLAARISRRNARNARVKEVFD 267
QY 1080 HPTLAGISAVVGDPLSYTLIPKSTHEGPEVQSYSGRLWFLDQDLVDGSLWILIPYAVRM 1139
Db 268 QPVLAIDLAIQGRSKPHNPILTAPVSGPVEQSFAGRLWFLDQINLGASWGLMPLAVRL 327
QY 1140 RGPVNVDAURLAALAEORHEILRTTFEDQDGVGOIVHEKSEEMKVITDLGSLDLP-- 1197
Db 328 RGPLHEALTITAILAEORHEILRTTFEERDGVGVVREHLTEKLVRIDV-PADLDGGY 386
QY 1198 FEVLNQEQTPPNSEAGWRATLLRLGDDHLITIVMHHIISDGWSIDVLRDLNQLYS 1257
Db 387 LQPLKQEQTSFDLASEGWRVSLRLRIANDHVLISVMHHIISDGWSIDLIRRELQQLYS 446
QY 1258 AALKDSKPLSALTPLIQYSDFAKWK--DQFIEQKOLNYWKKOLKDSKPAKIPTDFA 1315
Db 447 AALR-GYDQSQSLPLPIQYRDFSVNQKAEQVAEHEROLEYWTTLRADDSPAELTDL 505
QY 1316 RALLSGDAGCVHVITDGLYOSLRAFCHNEHTTSFVYLLAFAAAHYRLTAVEDAVIGT 1375
Db 506 RPTVLGSGNAVYQLAIDGSLYKLAFCRAYQTTSFAVLLAFAAAHYRLTGAEDATIGT 565
QY 1376 PIANRRNPELEDIGCFVNTQCMRINIDHHDTFGTLINOVKATTTAAAFENEDIPPRVYS 1435
Db 566 PIANRRNPELENLIGFVNTQCMRITVDRDDTFEILMQQVRSSTTAANFANQDVPFERIVS 625
QY 1436 ALOPGSRDLSSTPLAQILPAVHSQKDLGRFKQGLSVSPVSKAYTRFDMEFHLFOETDS 1495
Db 626 ALLPGSRDTSRNPVLQMEVLHSQKDLGMIQLEGVVGPEVPTAASRFDLEFHLFOEVR 685
QY 1496 LAGSYNFADELKMEIVENVVRVFEILNGLQSSRTPYSIILPTDGIYVLEKDLVNVK 1555
Db 686 LGGTVLFAFDLEFPDPIRGMTVTFQEVLHVLDQPHPTITSPLANGLAGLCNMGLLDIE 745
QY 1556 HVDYPRESSLADVFOTQSAVYDPSLAVDSSCLTYTELDROSDILAGWLRRRSPAETL 1615
Db 746 RSDYPRESSLVDVFEQVAACPDATAVIDSSRLTYQLDRQSDVAGWLRRRHMAETL 805
QY 1616 VAVFAPRSCETIVAFPGVLKANLAYLPDVRSPARKVDILSGSLGPTIVLIGHDTAPPD 1675
Db 806 VGVLA PRSCQITVAFGLILKANLAYLPDWNVPAARIEAILSAVEGHKLVLGSDVAAK 865
QY 1676 IEVTNVFVRI DALNDSNADGFEVIE---HDSKPSATSLAYVLYTSGTGPKGVMI 1732
Db 866 VRQDQVEMVRVDDTL-----GYDILDDYANSAPRSPSTSLAYVMTSGTGPKGVMI 919
QY 1733 HRVIRTV-TSGCIPNYPSETMAHMTATIAFGASVEIYSALLFGRTLVCVDYMTTLAR 1791
Db 920 HRAVLRLVKNVNSVILPMTPRVAHLSNLGFDISVQEVYALLNGTLVCIDYFTFLDSN 979
QY 1792 ALKDVPFRHVNAAHSVTSQDVPVRPRRLSRTLMFFFLVVYVTDSTAPDALDAQGLYOG 1851
Db 980 ALETVMQQRVAMLTIPALLKCLADIPSVLG-SLDVLFNVGDRPDRSDALAAQAL--- 1035
QY 1852 VOC--YNGYGPTENGV 1865
Db 1036 IRGCVYNAYGPTENGI 1051
```

RESULT 5

```
Q9L8H4
ID Q9L8H4 PRELIMINARY; PRT; 4247 AA.
AC Q9L8H4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Actinomycin synthetase III.
GN ACMC.
```

```
OS Streptomyces chrysomallus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1895;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11523;
RX MEDLINE=20243879; PubMed=10780924;
RA Schaeuwer F., Pfennig F., Grammel N., Keller U.;
RT "Construction and in vitro analysis of a new bi-modular polypeptide
RT synthetase for synthesis of N-methylated acyl peptides.";
RL Chem. Biol. 7:287-297(2000).
DR EMBL; AF204401; AAF42473.1; -.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR000379; Ser_estrs_site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00668; Condensation; 3.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00075; ACP_DOMAIN; 3.
DR PROSITE; PS00455; AMP_BINDING; 3.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
KW Phosphopantetheine.
SQ SEQUENCE 4247 AA; 462293 MW; 580A7F41522A0BC5 CRC64;

Query Match 14.8%; Score 2391; DB 2; Length 4247;
Best Local Similarity 26.7%; Pred. No. 2.3e-138;
Matches 888; Conservative 488; Mismatches 1282; Indels 664; Gaps 117;

QY 41 LQSSRIEAKPCTPFOLDIMDCNALDKQSAIGHA----VVDPTDIDISRALAWKEIVNQ 97
Db 1 MTSSGLIEDLPLSPMQEGLLFHSRYEQDGDVYAVQVHLVDLAGPLGRLREAVAGLVQR 60
QY 98 TPLARAFATSDSGKTSQVILKDSFVSNMCMSSSSP---DEVVRDEAAAAAAGPRCN- 153
Db 61 HPLNRAAFQVDSGRVTQQLISR-HVDLPWEEFDLSSMPAPEAAELSRIAAKCHGRFDP 119
QY 154 -----RFVLEDMQTKKQCVMTFSHALVDVTFQQRVLSRVFAAYKKHKDTHRPE-TPE 206
Db 120 AEPPLRLFTLVRTAE-EHRLILTHILLDGGWSTPLVRELTAALYADGPAALPRVTPY 178
QY 207 SSDATDTSQSVSVSMSCEDNAVSAHFQWTHLNDLNASVFFHLSDLHMLVP-NPTTTA- 264
Db 179 R-----QYLGWLAQQDRPAAEAAMREALGDLEQT-----LVAPVDPARAAL 220
QY 265 -EHRITFPLSOK-----ALNSAICRTALSILLSRYTHSDEALFGAVTEQSLP 311
Db 221 MPERITELAEETAALVEWARHGVTLNVLQAAGLVLSRRTGHHVDFVGVAAGRDP 280
QY 312 FDKHYLAD-----GTQYTVAPLRV-----HCQSNL-----RASDVMDAISYDDRLGLHLPFG 359
Db 281 ----QLAGVESVMVGLLITWVVRVRLDHAGSLLEAVRRLQDEQSRLTSHH----HLGIAR 332
QY 360 LRDIRVTGNGSAAQDFQ-----TVLLVTGDSHVNNINGINGFLOQITESHFF-- 405
Db 333 IQOLTGMSELFDTSLVFENYPMQDPAELPATGLRITPD-----LGRGDATHYPL 382
QY 406 --MPCNNRALLHCQMESSGALLVAYVDHNVDSLQTTRELQOFGHLIKCLQSPDLDS-- 461
Db 383 TLIAAPGRRLYLRLD-----YRGLDFRAEACALVER---LIRV----LDLAVA 424
QY 462 ----SMAEVLNMTFYDRAEI----ESWNSQPLEVDQTLIIHHEMLKA-VSHSPTKTAQAWD 513
Db 425 DAERPLGRVDLLTAERRELLHAGNGAGMPELEASL---PELFRARVAAAPDAVAVEYGE 481
QY 514 GDMTYSELONVSSRLAVHKSGLGRAQQAIIPIVFEKSKWVIASMLAVLKSNGNAFTLIDP 573
```

Db 482 EALRYELDERANLAHLAAHGV-GPERIVALALPRSDVLVAVLAVLAKAGAAAYPLDP 540
QY 574 NDPARTAQVOTRATVALTSKLHRETQKLVGR---CVVDDDELQSVS---ASDDFS 627
Db 541 EYPANRLAHWTDQPTVLVTT---TTEAKLPDRHPGCLLLDDPAVLADLSGRPAHPDVV 597
QY 628 SLTKSDLAYVFTSGTGPDKGIMI-----EHRAFSSCALKGASIGNSDTR 676
Db 598 ELHPDHP-AVIVTSGTGVPGKVMPPAGLLNLLQWHRA-----VGDEPQTR 645
QY 677 ALQGFTHAFACALLEIMTTLINGCCVCIPIPSDDDMNS--IPSFNRNRYNWMATPSYMG 734
Db 646 TAQFTAISFVDSQAQVLSVAFKTLVPIDEEYRRDAAARFAGLDDQVDELFAPNLVLE 705
QY 735 TFPEDV-----PGLATLVLGE--OMSSSVNAIW--APKLOLLNGYGOSSESSICFAS 784
Db 706 ALAEAAVETGRTLPOLRTVAQAGEALTLSRTVAFHRSAPGRRLHNYHGTFE-THVVTAH 764
QY 785 NMSTPNP-----MGRV--GAHSWIDPNDINRLVPIGAVGELVIESPGIARDYIVPPP 837
Db 765 ALGDDPEDWRLPAPIGRPIDNTHAY--RTRAVRLVPGVVGELYIAGAGLARGYLGRPA 822
QY 838 PEKSPFTDIPSWYPANTFPGAKLYKTGLARYASDGSIVCLGRIDSQVKIRGORVELG 897
Db 823 LTAERFVAD-----PYGLEP--GGMYRTGDLVRNPDGELEFCGRADHQVKVGRFIEPG 876
QY 898 AIETHLRQOMPDDLTIVVEATKRSQSANSLSLAFILGSSYFGNRPDSAHILHDHATKAI 957
Db 877 EIEKVLTDH--PDIAQAAV--VTRPHRPGDRLVAYVVGREAL--RP-----917
QY 958 NIKLEQV-----LPHSPISPYICMLPRATGKIDRRRLRIMGKIDLKQTOGAIV 1010
Db 918 ----EQVREFTRERLPEHMPVAAVQLERLPTPNKGLDRAAL-----956
QY 1011 QQAPAPIPVADTAAKLHS-----IMVQSLGIDPATVNVGATFELGGSNTIAKMW 1062
Db 957 ---PEPDFALAGGREARTQEQIVCDLFAQVIGL--PWVGVDDEFFELGHHLLATRLI 1011
QY 1063 NMAVS--GMDLKYNSIYQHPITLAGISAVVKGDPLSYTLIPKSTHEGPVEQSYQGRLWFL 1121
Db 1012 ARIRAAFSVELGLRTLFEARTAAVAHAHLDLAGPARTALTQHOLPDPAVPLSFAQRLWFL 1071
QY 1122 DQLDVGLWTLIPYVNRMPGVNDALRALALEQRHETLRTTFDQDQGVGOIVHEKL 1181
Db 1072 HKMEGPSATYNIPLALRLTGLNELALRAALEVDVTRHESLRTVFPEDVTPYQORVLDTA 1131
QY 1182 SEEMKVTDLCGSDLDPEVLNQEQTPPNLSSEAGWRATLLRGEDDHILTIVMHHITSD 1241
Db 1132 TIEL--VRTATTAEALTDHLRAARHPFDLAGEPPLTELFSAREHVLVLLVHHIAGD 1189
QY 1242 GWSIDVLRDLNQLYSAALKSDKDPLSALTPLPIQYSDFAKWQDOFIEQE-----KQ 1294
Db 1190 GWSLGPLASDLARAY--AARAEGRAPQWA--PLPVQVADYTLWQNELLDQNDPSLFPATQ 1246
QY 1295 LNWKKOLKD--SSPAKIPDTPARALLSGDAGCVHVTIDGELYQSRAFCEHNHTSFVV 1353
Db 1247 VAYWETETLAGLPQITLTPDRPFAVMTYRGDYLTVDIDPELHRLRTELARGSGASLEW 1306
QY 1354 LLAFAAAHYRLTAVEDAVIGTPIANRNRPELIEDIGCFVNTQCMRINIDHHDTFGLIN 1413
Db 1307 LQAGLAALIKRLGAGDDIPGSIAGRTDQALDOLGFFVNTLVLRITDITGDTFTQLLT 1366
QY 1414 QVATTAAAFENEDIPFERVVSALQPGSRDLSSTLAQLIFAVHSKQDLGRKFQGESV 1473
Db 1367 RVRETSLAAAHQDVPFEYLVEVLNP--TRTLAHPHPLFQIMLALQNAPE--GTFQLPGLTVD 1424
QY 1474 PVPKAYT-RFDMEFHLQ-----ETDSLKGSVNEADELFKMETVENVVRVFEILLNG 1526
Db 1425 VAPGRTGTAKFDLFFLSLAERRGADGEPQGITGAVEYSDDIYDAPTQVALFNRIHLDTA 1484
QY 1527 LQSSRTPVSLPLTDGTLVLEKLDVLRNVHVDYPRE---SSLADVOTQVSAVPDSLAVV 1583

Db 1485 TAQPEOPLSRIDL-----LTAQEHQHLDTWLDTAVEVGPDLPLPARFARQAATPQAVALI 1540
QY 1584 DSSCLRTYELDRQSDITLACGLRRRSPAEITVAVFAPRSCETIVAFEGVLKANLAYLPL 1643
Db 1541 AGDLSLYAELDARANRLAHALLREGAGPDRBLVALALPRTAELVALLVLTGCAAYLPL 1600
QY 1644 DVRSARVODIILSGLSGPTIVLIG-----HDTAPPDIEVTNVFVRIRDALND--SNADG 1697
Db 1601 DPDHPAARLSHVL--GOARPALLLTDTREQLHPADADTRRLALDLSAEVALLADCPOTDP 1659
QY 1698 FEVIEDHSTKPSATSLAYVLYTSSGTRCPKGMLEHVRVIRTYS--GCIPNYPSEPRMA 1755
Db 1660 AE-----EGVTPAGSAAYIVYTGSTGRPKGVVPHSALNVFVTAMRRQAPLQOE--RLL 1714
QY 1756 HMTIADFAGASYEYISALLFRTLVCVDYMTTLDARALKDVFVFEHVNAAASHVTS-----1810
Db 1715 AVTTVAFDIAALELYHPLLSSAAVVLAPKAVQPSAVLDLIARHGVTTVQGTPLMQOLL 1774
QY 1811 -----SSQDVLVRPRRLSRTLMMFFFLVLTVDSTAPDALDAQGLYQVOCY 1855
Db 1775 VGHDAEALRGLRMLVGGEALPLSLAEALR-----ALTDOLV-----1810
QY 1856 NGYPTENGVMSTIYPIDSTESFINGVPIGRALNNSGAYVVDPEOQLVGVGMGELVVTG 1915
Db 1811 NLGYTETTITWSTAAEL---AGGTGAAPIGRPIANTRVYVLDLQLOPAPVGVVGLYIAG 1867
QY 1916 DGLARYSDK--ALDENRFVHITVN--DOTVKAARTGDRVRYRIGDGLIEFFGRMDTQFKIR 1973
Db 1868 AGLARYGLDRPALTAERFPADPYGLEPGRMYRTGDLVRN--PDGELEFFVRADHOKVR 1926
QY 1974 GNRIESAEIPALLRSSVRDAAVVOONEDQAPELGLGVFVADHDHSENKGSANOVEG 2033
Db 1927 GFRIEPEIEKVLTDHPDIAQAAVVRREDQGDARLVAYVYVVGTSADARQEVEQDQDE 1986
QY 2034 WQHEFSGMYSDIGEIDPSTIGSDEKWTSMYDSQIDDFDEMHEWLGETT---RTLHNR 2090
Db 1987 WQOLYDS--VITASE---TAFGENFASWNSYDGLPIPLDOMREWRDTTVERIRGLNRR 2042
QY 2091 SLGNVLEIGTSGMILFNLDLSRLSEYVGLFSPSSAAAFVFNKATESIPSLAGKAKVQVGT 2150
Db 2043 ----VLEIGVGTGLLSRLAPHCCEYWGTDFTPTVIADLRGHVEADPELAARVQLRTQPA 2098
QY 2151 TDIGQVDDLHPDLVNLNSVIQYFSPSEYLAIEADTLIH-LPNVQRIFFGVDVRSQATNEHF 2209
Db 2099 HFDQDLPHGHFDVNLNSVQYFPPNAGYLEQVLDHALRILAPGTVFIGDIRNPLRLTF 2158
QY 2210 LAARATHTLTKNATKDD--VROKMAELEDMEELLVEPAFFTSLKDRFPGLVEHVELTPK 2267
Db 2159 --TTAVHTARATPDADHAAVRRRAVEHALVEKELLIDPEYFTALGHVHPDLA-GVDIRLK 2215
QY 2268 NMEAVNELSAYRAAVVHVHVRGSLGDELVLPEVKDDWIDFQANQLNOKSLGDLKSSDAAI 2327
Db 2216 PGTVENELTRYDATLHKAGATHPLTGPPEP--PWSGHADPDALAEALARQLDSEPAR 2274
QY 2328 MAYSKIPFETATERQOVVASLNSIDENWQLSTRSSAEGSSLSVDP---IFRIAGEAGF 2384
Db 2275 LRVGTGPNPLARELALQHALDT-----ATAPIAPTADAQGPQDPAALHRLGRHQY 2329
QY 2385 RVEVSSAROWSON--GALDAVFHCCS--QGRTLVNPETDHLHRLGSDLLTRNRLQORNR 2441
Db 2330 ----WTAATWNTHRPDAVDLVVALSDLGAT-----PTGTYSAGTSTPATPLSSLTNP 2381
QY 2442 IA-----IEVRERLSLLPSYMIPIINSIVLDKMPNLNANGVKDRKLSRRKAVVPKQ 2493
Db 2382 AAGRGTSALLTTUREHARTHLDPYMQPSALVPLDRLPLTANGKLDL-----2427
QY 2494 TAAPLTPPLISEVEVILCEEATEVFGMKVDITDHFNFNLGCHSLLATKLSRIDORLKVRI 2553
Db 2428 AALPAPDFTLA-----2438
QY 2554 TVKDVEDHPVFDLASVIRGLQOPVSDGQDRSAHMAPRTETAILCDEFAKVLGF 2613
Db 2439 -----GTGRE-----PRTPOQIVCDLFTQVLGL 2462

```
QY 2614 -QVGTIDNPFGLGHSLSMATKLAIRIGHRLDITTSVKQVDFHPVLFQFLAIALDNLVQSKT 2672
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2463 PRVGIDDDFELGGHSLATRLIAQIRALFGVELESLRSLFEGTTPAANAARLDTA----- 2517
QY 2673 NEIVGGRMAEYSPQLLET--EDDEEFMASEIKPOLEQLQEIIDIIYPSTQOKAPLFPH 2730
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2518 -----GPGRLALTVOERPATM-----PLSFAQRRLWFTH 2546
QY 2731 TTARPRFPVFFIDFPSTSEPDAAAGLTKACESLVNHLDFRTVFAEASGELYQVVL---- 2786
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2547 KMEGFSATNIPDLAULSGLDRDALRGALTDLVSRHESLRVFFPEVDGTPQRLLTPEA 2606
QY 2787 -----SCLDLPQIVETEDNINTAINTEFLDEFAKEPVRGLHPLIRFTIIKOTSM 2836
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2607 AAPRLTVTPTSEADLP-----DALEAAARYAFD-LAEQP-----PL-RTELFELSARE 2652
QY 2837 RYIMRISHALY-DGSLSEHVVRKHLMYNGRSLPPHQPSSRYMYTADRGSGHGFWRDV 2895
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2653 HVLLLVVHHIAGDGSGLASPLASLARAYAAR-----AEGRAPO---WAPL 2694
QY 2896 -IQNTPMTILSDTVDVGDND-----ATKALHLSKIVNIPSOV----- 2932
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2695 PQVADYTLWQNELLDQNDPSLFATQVAYWTETLAGLPDQITLPTDRPRPAVMTYRGD 2754
QY 2933 -----LRGSSNIITQATVFNACALVLSRESKDVVFGRIVSQROGL 2975
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2755 YLTVDDIDPELHRLTELARGSGASLFW--VLQAGLAALLKRLGAGDDIPLGSPAGRTDQ 2812
QY 2976 PVEYQDVGPCPNVAVRAHISS--DYNQLLHDIOQOYLLSLPHETIGFSDLRKNCNDWP 3034
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2813 ALD--DLIGFFNTLVLRDTTGTGDEFTTOLLTRVRETSIAAYAHQDVPEYELVE----- 2864
QY 3035 EAITNFSCCITYHNFEPHESQEQOVEMGVLTFRVNIEMDEPLVDLAIAGEVEPDGA- 3093
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2865 --VLNPTRLAH-----IAKTOLF 3106
QY 3094 ----GLKVTV-----IAKTOLF 3106
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2893 QLPGLTVDVAPGRTGTAKFDLF 2914

RESULT 6
O07944 PRELIMINARY; PRT: 4848 AA.
AC O07944;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Pristinamycin I synthase 3 and 4;
GN SNDBE.
OS Streptomyces pristinaespiralis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=38300;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp92;
RA de Crecy-Lagard V.A., Saurin W., Thibaut D., Gil P., Naudin L.,
RA Crouzet J., Blanc V.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp92;
RX MEDLINE=97158664; PubMed=9006024;
RA de Crecy-Lagard V., Blanc V., Gil P., Naudin L., Lorenzon S.,
RA Famechon A., Bamas-Jacques N., Crouzet J., Thibaut D.;
RT "Pristinamycin I biosynthesis in Streptomyces pristinaespiralis:
RT molecular characterization of the first two structural peptide
RT synthetase genes.";
RL J. Bacteriol. 179:705-713(1997).
DR EMBL; Y11548; CAA72312.1;
DR EMBL; X98690; CAA67249.1; -;
```

```
DR HSP; P14687; IAMU.
DR InterPro: IPR002106; AACRNA_ligaseII.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR002336; ErythrurIn.
DR InterPro: IPR001601; Methyltransf.
DR InterPro: IPR003880; Pplantne_attach.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR000379; Ser_estrs.site.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 4.
DR Pfam: PF00668; Condensation; 5.
DR Pfam: PF00550; pp-binding; 4.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PR00154; AMPBINDING.
DR PRINTS: PR00611; ERYTHCRUORIN.
DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE: PS50075; ACP_DOMAIN; 4.
DR PROSITE: PS00455; AMP-BINDING; 4.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_4.
DR Phosphopantetheine.
KW SEQUENCE 4848 AA; 522081 MW; E17591617A2B9A0E CRC64;

Query Match 14.1%; Score 2275.5; DB 2; Length 4848;
Best Local Similarity 26.7%; Pred. No. 4.5e-131;
Matches 877; Conservative 474; Mismatches 1280; Indels 659; Gaps 125;

QY 72 GHAVYDV-----PTDIDISRFALAWKEIVNOTPALRAFAFSDSGTSQVILKDSFV 123
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 31 GPDVYTVQMVFEURGLPDEDDGURAAA--ALLRRHPLNLRAGFWQQGVPRVQV----- 82
QY 124 FSWMCWSSSSSDEV-----VRDEAAAAA--PNEVPLPWHTRDFALGESDRERAAYVATDRAERFDGAPPLIRFAL 157
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 83 -----PNEVPLPWHTRDFALGESDRERAAYVATDRAERFDGAPPLIRFAL 131
QY 158 LDMQTKKQVWTFSHALVDVTFQORVLSRVFAAYKHEKDT--HRPETPESDADTDTDS 215
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 132 L-GLAADHKLVLTHHLLLDGWSMPLRLFTLYGQRCDDAGMPVTPYRAYLAHLAG 190
QY 216 QSVSVVSMSCEDNAVSAATHFWOHLNDL-----NASVFPHLSDLHWVP-----NPT 261
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 191 R-----DDAARAA--WRTALADLEPSLVAGAGAGAGAGAGSALPGQIWHEDAA 239
QY 262 TTAHRITPPLSOKALSNAICR-----TALSILSRYSRTHSDEALFGAYTEQ- 308
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 240 TTA-----GLTALARS--CNITLQHPGAKRLGTAARPOLGR-----DDVVFQATVAHR 285
QY 309 --SLPDKHYLDGTYQTVAPLRVHCQSNLRASDVMDAISSYDDRLGHLAPFGLDIRWT 366
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 286 PPEIGIESTI--GMFINTLPVRVVRPAETLGLDLGRVQREQAALIEHRHLSLTDIRST 343
QY 367 GNGSAACDFQTVLL-----VTDGSHVNNINGFLOQITTESSHFMPCNNRALLLHC 417
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 344 TGTGEL--FDTVVVFENYPLDPAVLRARAGRLAGF--EVSATHY-PLSLAI----- 393
QY 418 QMESSCALLVAYYDH--NVIDSLQTRLLQQFGLHLKLCQ-----SPLDSSMAEYN 467
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 394 ----PGESLRFRLDHRGDVLDEAGARLLRLERLTDIAEHGADLPVGRGLDLSAAERH 449
QY 468 LMTYDRAELESWNSQPLEVQDTLIHHEMLKAVSHSPTKTAIQAWGDWTDVSELDNVSSR 527
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 450 QV-----LEEFNDTGLPAEDATLAALFEAQAAPTDPTTALLVGGRSRLTYAEUNARNR 502
QY 528 LAVHIKSLGLRAQQAIIPIVFEKSKWVIASMLAVLKSNGAFTLIDPNDPAPARTAQVVTQ 587
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 503 LARHLVTLGAGPEQ-IVAVKLSRLDLYVALLAVLKTGAAYLPVDYTAIPAERLAPMDDA 561
QY 588 RATVALTSKLHRETQVKLVGRVVDDELQSVASDDFSSLSKSDQ-----AYVIFTS 642
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 562 RPAVLTG---DTGQDLG-----YDD-----TDLTDAORTAPLLPAHPAYVIYTS 605
QY 643 GSTGDPKGLMIEHRAFFSSCKAFGASLGINSDFRALQFGTHAFGACLELMTWTLINGCV 702
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```



```
Db 4217 LCNLAQAQIAQFGLHSDSRILOFASLSDACISEILMSLGSQGLT-----YLGSKDSLMPG 4272
QY 1791 ----RALKDVFREHYNAAASHVTSQSDQPLVRPRRLSRTLMFFFLVTDVSTAPDALDAQ 1846
Db 4273 TPLIERLRY-----AITHILPSS--ALAVLHKVLTTLTIIVAGEACAVELIKOW 4323
QY 1847 GLYQGVQVNGYKPTGPTNGYMSIYIPIDSTESFNGVPIGRALNNGAYVDPPEQQVLVGIG 1906
Db 4324 S--AGRNFFNAYGPTGEGSCAT---LAKSALTQKLPVGPITANQVYILDSQLQVPVIG 4378
QY 1907 VMGLVVTGDLGARGYSDK-ALDENRFVHTVNDQTVKAYRGCDRVRYRIGDGLIEFFGR 1965
Db 4379 VPGLHAGVGVARGVNLNPELTQKRFISNPFGPS--KLYKTGDLGRY-LADGTIYELGR 4435
QY 1966 MTQFKIRGNRIEASIEAALLRDSVRDAAVVLOQNEQDQAPFILGVLVADHDHSENDKG 2025
Db 4436 IDNOVKIRGFRIELGEIEAVASOHPLVQESVVIAREIDPCDKRLVAYLVPAL-----QG 4489
QY 2026 Q-----SANQVEGWQDHFEBSGMYSDIGEIDPSTIGSDFKGWTSMYDGGSIDFDEMH 2076
Db 4490 QVLPEOLAQWQGEYVSDWQKLYEQSYSQOQTPDDPTF--NISGWNSSYTGKRAIPDSEMR 4547
QY 2077 EMIGETTRTLHDNRSLG--NVLEIGTSGMILNLDLSRLSESVGLPEPSRAAFVFNKAT 2133
Db 4548 EWVESTVSR-----LCKGQFVLEICGSGLLFRVAPHCQYEWGADYSATIRNLERLC 4603
QY 2134 ESIPSLAGKAKVQVGTATDIGQVDDLHPDLVNLNSVIQYFSPSEYLAETAD-TLIHLPNV 2192
Db 4604 GETIQGLE-NVRLHKTADFEGIPQCAFDTVVNSVYVFPSPDYLLQVLEGAMTAIASQ 4662
QY 2193 ORIFFGDSVQATNEHFLAARAIHTLGNKATKDDVRQKMAELEDMEELLVEPAFTSLK 2252
Db 4663 KTFVGDVRSPLPLLPYHAAQLARAESDKTVEQWQQQVHOTVAAEEELLIDPRFFIALQ 4722
QY 2253 DRPPGLVEHVEILPKMNEAVNELSAYRAAVHVRGSLGDELVLPEK-DDWIDFQANOL 2311
Db 4723 QRP-QITWVEIQPKKGHQNELTQRYDVTLH-----LVLMGKSSLVKLATYQL 4773
QY 2312 NOKSLGDLKSSDAAIMAYSKFIPEITAFERQVVASLNSNIDWQ-----LSTI-----RS 2362
Db 4774 SWTQLQDLKLVHPWELGIRDVFNQRLQARQV-----LGMWENPPAASTVGELRQRL 4826
QY 2363 SAEGLSSLVDPIDFRITRAGAGRYEYSSARQW---SQNGALDAVFHCCS-----2409
Db 4827 AAQPMVGINPEQWIQGEELGYTVHLSS---WMESSQDGCDFVVLGRNSSTPVSIDIQSY 4882
QY 2410 ---QGRTLVNF-P-TDHLRGSDLLTNRLQRLQNRRIAREVRRLRSLPSYMIPIVNI 2465
Db 4883 SFWEHETITTKPWTY-----TNNPLYGLVKLVQVQVREFIOOKLPNMIPOAFVL 4934
QY 2466 LDKMPLNANGKVDKRLSRAKVVVVKQOATAA-PLPT---FPISEVEVILCEEATEVFGM- 2520
Db 4935 LNALPLTPNGKVDRRAL-----PTPDTASRNLTGTFVSPRTPIEAQMAQWSEVLGLE 4987
QY 2521 KYDITDHFENLGHSLATLKRIDQRLKVRITVXDFVDPHVPFADLASVIRGGLGLOQ 2579
Db 4988 RIGVKDNFFELGHSLLATQVTSRSDIFSVLSLQNLLEYPTIVNLAQMI-EVLGVSQ 5045
```

RESULT 9

```
Q01135 PRELIMINARY; PRT; 5157 AA.
AC Q01135;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Peptide synthetase.
GN PESA.
OS Metarhizium anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
OX NCBI_TaxID=5530;
```

```
RN SEQUENCE FROM N.A.
RP STRAIN=ME1;
RX MEDLINE=97082966; PubMed=8964498;
RA Bailey A.M., Kershaw M.J., Hunt B.A., Patterson I.C., Charnley A.K.,
RA Reynolds S.E., Clarkson J.M.;
RT "Cloning and sequence analysis of an intron containing domain of a
RT peptide synthetase from the entomopathogenic fungus Metarhizium
RT anisopliae.";
RL Gene 173:195-197(1996).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ME1;
RA Bailey A.M., Reynolds S.E., Charnley A.K., Clarkson J.M.;
RT "Evidence for multiple peptide synthetases from Metarhizium
RT anisopliae.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; X89442; CAA61605.1; -
DR HSSP; P14687; YAMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Pantane_attach.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; Condensation; 7.
DR Pfam; PF00550; pp-binding; 4.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
DR PROSITE; PS00455; AMP_BINDING; 4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Phosphopantetheine.
SQ SEQUENCE 5157 AA; 573954 MW; 1038242BA3143868 CRC64;

Query Match 12.5%; Score 2018.5; DB 3; Length 5157;
Best Local Similarity 23.9%; Pred. No. 4.9e-115;
Matches 795; Conservative 536; Mismatches 1278; Indels 711; Gaps 127;

QY 41 LDSSRIEAKPCTPPFOLDMIDCNALDKQSAIGHAVYDVPTDIDISRFALAKWEIVNQTPA 100
Db 2196 VKDCIEDIYPCPTQEGLEMAISTRPRAYVARVARKFKIPSTIDLNRFRNAWQLISLEPI 2255
QY 101 LRAFAFTSDSGKTSQVILKDSFVSWCMWSSSSSPDEVV-RDEAAAAAAGPCNRFVILE 159
Db 2256 LRTRVLITNTNSV-QVVLREQ-----NPNKEGLTLESLKDNGLPIEYGLLHRLCIID 2309
QY 160 DMQTKKCOLWTFESHALVDVTFQORVLSRVFAAYKHEKDTHRPETPSSDATDTSQSVS 219
Db 2310 GL-GQDSYFIWSVHHALYDGSQNLFFERVYKQLY-----MGDAISKSPSYNR 2355
QY 220 VYSMCEDNAVSATHTWOTH-LNDLNASVFPHLSDHLMVNPNT--TTAER--RITPLSQ 274
Db 2356 FIHFLAQEDTKSANKFWKSQLREQAPTSPP-----VLPSPYKPRADHVQTLRLPLSR 2409
QY 275 K---ALSNSAICRTALSILLSRYTHSDEALFG-AVTEQSLPFDKHYLADGYQTVAPLVR 330
Db 2410 QKHSTITSATLKAANAMVVSVQSRSDDIILALTLSGRTAPVPGIMKMTGTTITTVPLRI 2469
QY 331 HQSNLRASDVMAISSYDDRLGHLAPF---GLURDTRNGDNGSAC-----DFQT 378
Db 2470 ELPS--RNTTVYKYLELVQNAQEMMPYEHVGTQSLRR-----LCLEAEPDYELDKH 2520
QY 379 VLLVTDGSHVNGINGFLO---QITESSHFMPCNNRALLLHCOMESSGALLVAYYDHNVI 435
Db 2521 LLVIQSLGEV---DGLFLELLPLETDEDDF---NTYALIVQCSVEENVCIEARYDRDVI 2573
QY 436 DSLQTTTRLQQFGHLIKCIQSPDLSSMAEVENLMTEDYRAEIESWSQPLEVQDTLTHHE 495
Db 2574 ETQGMERMLWQFOHVLRLQNLNDESHSSTMEGVDLVSPQDLQDLVWNSQSHESMDICIQV 2633
QY 496 MLKANVSHSPKTK-AIOAWDGDWTYSLELDNVSSRLAVHIIKSLGRQAQAIIPVYFEKSKWV 554
Db 2634 IEFQMRSRVPDAPICAWDAGELSYQELDRLSNGLATHLVDLGV-GPEVMVPICFDKSAWT 2692
QY 555 IASMLAVLKSNGNAFTLIDPNDPPARTAAQVVVQTRATVALTSKLHRETQVOKLVGRCVVDD 614
```


Db 1751 SLSYGRKDTQIKHQRIEIGEIEHHV-LHCTKAVEVTVDVAVVPGEKNKSLVAFV-- 1807
QY 936 SSYFGRNPSDAHILD--HDATKAINIKLEQVLPRESIPSYFCMLPRTATGKIDRRRL 993
Db 1808 -----RPSNGTSTPQFDNDPAIINELANSLPAYMIPTMYIQVPSIPTASGKTRKQL 1861
QY 994 RIMGKIIDKOTQGAIVOQAPAPVPFADTAALKHSIWQSLGIDPA-TVNVGAPFFELG 1052
Db 1862 REMGTAMASSHAARHWHQNRPPV---TDMKHHVKLWARVLTLENAGEISLDDSFIRLG 1918
QY 1053 GNSITAIKWNAMARSVGMDLKVSNIOYHPTLAG-----ISAVKGD-----LSVTLIPKSTH 1105
Db 1919 GDSIAAMKLVSIAAKAGLGLTVAQIFRHKLEDDQARHVTLTQGGPAPAFSLPDDSPD 1978
QY 1106 -----EGPVEQSY-----QGRLEWLDQDVGSLWYLIPIYAVMRGPVNDV 1146
Db 1979 VKALQADARAYAIEASSIEDVYPCPLQEGLLSLSSKPSYNTYTLQHVLPPTVDIQ 2038
QY 1147 ALRRALAEQRHETLRT-----FEDQDGVQIVHEKLSSEMKVVIDLGGSDLPDF 1198
Db 2039 QLRSAW-----EETIRTTDLTRIVLHPRYGL-VQVV---VKKEIQWHEPANADV--- 2085
QY 1199 EVLNOEQTTPFNLSSEACWRATLLRLGEDDHILITVMHHIISDGSIDVLRDLNQLYSA 1258
Db 2086 -YIETDKQVQWVLGSLVRYALSPDPTGSARKFIWTHALADGWTLDLILRKVKLAYST 2144
QY 1259 ALKDKDPLSALTPLPIQYDFAKW---QKDOFTEQKQLNWKKQKLDKSDSPAKIPTDF 1314
Db 2145 -----LHTVSPVS-EFRSVKVIITRNDMEV-----YKSTLGGYHSTTFPVC 2189
QY 1315 -ARPALLSGDAGCVHWTIDGELYQSLAFACNEHNTTSFVLLAAFAAHRYLTAVEDAVI 1373
Db 2190 VSTVAIEDSEVGQKH-----ELPRNITLSAHLST-----LLRAAWAIVQSNYSNTSDVVF 2240
QY 1374 GPPIANRR--PELEDIIGCFVNTOCMKNIDHDTFGTLINQKATTAFAFENEDIPE 1431
Db 2241 GEVFSGRSASVPFIEAIGPTMATLPRVKIDDSAREMLDRLTTTOMIPHQOLGLO 2300
QY 1432 RV-----YSALQPSKDLSSSTPLAOLIFAVHSOKDLGRFKFOGLSVVPV 1476
Db 2301 RISQINTDQACASFOTLLVQPPASTHNGQEPLSFSGSPDYRLATYAL-GIECTPA- 2358
QY 1477 SKAYTRFDMEFHLFOETSLKSVNFADLKFMEVENVVRVFFELRNGLOSSRTPVSI 1536
Db 2359 SDGY-----SFCRARFDSRVLSAQVAERMAOLGHVVSOLVAVTASPSS 2404
QY 1537 LPLTDGIYV-----LEKLDVNLKHYDYPRESS---LADVOTQVSAYPDSLAVYDSSCR 1588
Db 2405 TLVSIVLNTPODLEKLANWNAEVLGEQKHSMLLHQVFRKKALAAPQATAISSWDGE 2464
QY 1589 LTYTDLRQSDILAGLRRRSMPA---ETLVAFAPRSCETIVAFFGVLKANLAYLPLDV 1645
Db 2465 CSYAQLEKLSDALAAMLTDGIGLDQOLVPLCFERSMWVVMVAMVLTGAGIVPLDP 2524
QY 1646 RSPSARVQDILS--GLSGPTIVLIGHDAPPDIEVTVNVEFVRIRDALNDSNADGFEVIEH 1703
Db 2525 AHPPSRHERILAKVIGG--CILVSPQVQAQROFGEGWTTW-----VSEASAAVPSI-H 2576
QY 1704 DSTKPSATSLA--VYLYTSGSTGRPKGVMIHRVIRTV-TSGCIPNVPSETRMAHMATI 1760
Db 2577 AFDPTVTHLAVCWILFTSGSTGCEPKGYILEHGATCSYKLLGKTLGDKETRMHLFSAY 2636
QY 1761 AFDGASYEYISALLGRTLVCV-----DYMTTLDARALKDVFFREHVNAAASHVT 1809
Db 2637 AFDIATFEIIGLMSG-GCICIPSDAERLERLPQECTIF-----AVNTAILTP 2683
QY 1810 S-----SSQDPLRVPRRLSRLTMFFLVVTDSTAPDADAGLYQGVO-CYNGYGPTE 1863
Db 2684 SVARLYTPNDIP-----TLRSLCLA--GEAPNKQDISTWQHRRIPFLFNCYGPAAE 2731
QY 1864 GWMSTIYIPDSTFESFINGVPIGRALNNSGAYVVDPE--QQLVIGVMGELVVTGGLARG 1921
Db 2732 ACLATNIRGNPDARSATRIGR-LRGVPLWITAPGNCRKLAPIGAVGELLIEGTLARG 2790

QY 1922 YSDKALDENRFRVHIT-----VNDQTVKAYRTGDRVRYRIGDGLIEFFGRMDTQF 1970
Db 2791 Y---WQLKPMHLHSRTSGCCCKAQIGSDPADVDSTGPEIVKYKDEDDGVVYEGRKDNQV 2847
QY 1971 KIRGNRIESAETEAALLR-----DSSVRDAAVVL-----QONEDQAPEILG 2011
Db 2848 KIRQORTELEIYHLSQCFPTAAEVVVEVATSERDLASVTILVAFVKRETRDSSEKVA 2907
QY 2012 FVVAHDHSENDKQOSANQVEGMOHFEESGMYSDIGEIDPS---TIGSDFKGWTSMYDG 2067
Db 2908 GIFALPSKLEHEINRRPL-----YMPAVFVSPEIPKPTATDKTDROKRELASVAT 2961
QY 2068 SQIDFDE-----MHE-----WL-----GETTRLHDN-----RSLGNV 2095
Db 2962 RAVDAPHQHPORLPSTVMEETLRDLWLKLVIPVQRTAIGLDSNFFRLGGDSIAALKLVGQA 3021
QY 2096 LEIG-----TGSGMI-----TGSGMI-----LFNLDLSRESV 2117
Db 3022 QOAGIELSKDIFLNPKLVDLAACCTDRRCVKEGSRMVAKHATISRFSLPLINASISIV 3081
QY 2118 -----GLEPSRSAAAF-VNKATESIPSACK-----AKVOGTATDIGOVDL--- 2159
Db 3082 DEVANACGIPPLVEDVPTCTPMQEGLSLSSRNPGTVVSQIAIELAPDV--LVLDLFKLA 3139
QY 2160 -----HPDLVVLNSVIQ---YFSSSEYLAETIADTLIHLPNVQRIFFGD 2199
Db 3140 WQOTVSTWPIILSRIIQHPKGLGFLQAVLKEDVTWNNSTDLEVEDSSTP---MGFGS 3195
QY 2200 VRSQ-----ATNEHFLAARATH-----TGSGMI-----LFNLDLSRESV 2216
Db 3196 ELSRHALVMDNSGKHIFVYTVHHSIYDHTVLRLLDDVDYDNYKGNRKDFQPTYSFVRS 3255
QY 2217 -----TLGKNATKDD---VROKMAELEDMEELLVEPA----- 2246
Db 3256 VISMKSSSEEFWRNACKDEGSSIFPQRSLSRESCEDTIVQESQLCTATGTVMANVL 3315
QY 2247 -----FFTSCLKDR-----FPGILVEHVEILPKNM 2269
Db 3316 HAAMAVSSWHVGNQSIQVFTVLSGRTPAVLGIENIAGPTIATAPPFVIDPSETISNFS 3375
QY 2270 EAVNELSAYRYAAVHVVRGSLGDELVLVPEKDDWDIDFO-----ANQLNQSLGDLKS 2322
Db 3376 PAIOGOMA---AVIHAOLGIGQIRLSLSSACELACNFOTLFAVQEGRAMVGNLSGKLL-- 3430
QY 2323 SDAATMAYSKIPFEITAF---EROVVASLNSN-IDEMOLSTI-----RSSABG 2366
Db 3431 -DVNTFSMRTYALTDCFLDTGTFHVKASDFDSRVVDOWRMESILRQFGAVAQOLATKABG 3489
QY 2367 DSSLVDPDIFRIAGBAGFRVYSSARQWSONGALD--AVFHHCCSQ-----GRPLV-- 2415
Db 3490 GELVS---SIETLNEQGWEL---LRRMNSHRTKQWAVFPEDECKPSPIGAIGELLIEG 3542
QY 2416 -NFTPDHHLRGSDLLTNRLPRLQRLNRIAEVRER---LRSLLPSWIFSNIVLVDKMLP 2471
Db 3543 PDFF-----SKYLEDGARGVRPMDRNGHKTVLLTGLILVAFDQNGNSIHIGQKRT 3595
QY 2472 NANGKVDRKELSRRAKVPKQOTAAPLPTFPFISEVEVILCEATE---VFGMKVDIOTH 2527
Db 3596 TISFKGORIDVSO-----IERHITSFLAGTEAVVEAIPAENSQSVALVFLHRPELADR 3651
QY 2528 PFNLGGHSLATKLISRIDQRLKRVTKVDHPFADLASVIRQGLGLOQPVSDGQGO 2587
Db 3652 GDNKSRPAICWSKQYDIEKNLSV-----VFPDMVPTL-----YIDMEAM 3691
QY 2588 DRSAH-----MAPRTETEAILCDE-----FAKVLGFO---V 2615
Db 3692 PRTHGIDIDRSQLOTGLSLFPAEKVAILRASQRKRPVAMQAIIRGLWASLIGAKEDTF 3751
QY 2616 GJTDFEELGGHSLMATKLAVRIGHRLDTTVSVKOVDFHPVLFQALADNLNVOSKTNEI 2675
Db 3752 HLDDOFFFKSGGDSIGVIKL-VGEARKNTALAAADIFQPKLESIA-----VRATENTL 3804

Db 380 QLLSRVRLTQAYAHDPDFESLVKVLQP--ORDLSHTP-----LFQVMEVLQNAPI 430
QY 386 SHYNNINGFLOQITESSHFMCNRA---LLHCOMESSGALLVAYDHNVIDSLQTT 441
Db 431 SEVE-----LADLTISS--LPAQSATAFDLTLSMONTATGLVSWYENADLFDASTIE 482
QY 442 RLLOQFHLIK-CLOSPDLSSMAEVLNLTXYDRAE-IESWNSQPLEV-ODTLIHEMLK 498
Db 483 RMRGHEFTLLEGIVANPQE--QISQPLLTETEBQOQLLKWNTGDYDQDKCIHLLFEE 540
QY 499 AVSHSPKTAIQAWDGDWYSELNDVSSRLAVHIKSLGLRAQAQAIIPVFEAKSWIASM 558
Db 541 QVELTDAVAVYENQHLTYHELNCRAQLAHYKLSGVKA-DVLVGICVERSELMVGL 599
QY 559 LAVLKSGNAFTLIDPNDPAPARTAAQVVTQTRATVALTSKLHRETVQKLVGR-----C 609
Db 600 LGILKAGGAYPLASDYPFERLRLMLEDAQVSVLLTQ-----QKLIDRLPHTAQRVC 652
QY 610 VVDDDELQSVASDDFSSLSKSDIAYVIFTSGTDPKGINIEHRAFSSCALKFGASL 669
Db 653 LDADQWVISOLS-QDNLIPETQATNLAYVIYTSGSTQPKGVLIHAGHLLNLVFWHOFTE 711
QY 670 GINSDFRALQFHAFGACLELMTTLINGGCVCIPSDDRMNSIPSFNRRNVNMMMA- 728
Db 712 KITTLDKATQACTADAAWELWPLYTAGASIIYLVKPEILSPVDL-----QDWLESK 765
QY 729 --TPSYMGTFSPBDV-----PGLATVLVGEOMSSSVNAIWAPEKLOLNGYGOSE-- 776
Db 766 KITISLPMPABOLLSLEWPESTTLRLMTLGTGDKLHRYPSGLL--PFQVNNYGPTEW 823
QY 777 --SSSTCFASNSTENNNGRAVG-----AHSWIDPDNDINRLVPGVAGELVIESPGIA 829
Db 824 VVSTGLVVSNN--GRDNNISPPGIRPIANVEIYILD--SYLQPVVPVGPCELHIGGAGLA 879
QY 830 RDYIVPPPEKSPFFTDIPSWP-----ANTFPDQAKLYRTGDILARYASDGS 876
Db 880 RGYLNPOLTOEKFIN-----PFKRSRGAGEORSRGETF--NSNRLYKIGDILARYLPDGT 933
QY 877 IVCLGRIDSQVIRGORVELGAIETHLRQOMPDDLTIVVEATKRSOSANSTSLIAFLGS 936
Db 934 IEYLGRIDNQVKIRGPRIELSEAEVLSSHSDVOLSCVI---AREDTGDKRLVAVV-- 988
QY 937 SYFGNRPDAHILHDHATKAINIKLQV---LPHRSIPSYFICMLELRTATGKIDRRR 992
Db 989 -----AH--QDCKPAIS--ELRQILKAKLPDYMVPFAFVLKSLPTNGKVDVRA 1035
QY 993 LRIMGKDILDKOTQGAIVQQAIPVPFADTAAKLHSIWVQSIGIDPATVNVGATPFELG 1052
Db 1036 L---PKPDLDT---TILEKYVAPRTPPIEMLALL--NAQVLKLE--LVGIYDNEFELG 1083
QY 1053 GNSITAUKMNNMARS-VGMDLKVSNYIYQHTLAGISAVV-----KGDPLSYTLIPKSTHE 1106
Db 1084 GHSLLATQVRSRITWFKVELPLRELFAASATVVELAQSIGRLQQOQDLELSTPFIQRAEN 1143
QY 1107 GPVEOSYSGRWLWFLDQVGLWLIPIYAVRMGPVNDALRRALAAEQRHETLRTTF 1166
Db 1144 AQLLSTAQORLWFLDQFPNAGALYNIPALRLVGTNLQVLALEQSLHEIARHEALRTNF 1203
QY 1167 EDQDGVGVQIVHEKLSSEMKVIDL-----CGSDLOPFEVLNQEQTTFFNLSSEAGWRATL 1221
Db 1204 VTVNGKPSQIQTQNTWTVSVVGLKHLSTTEQEIATQQLVROQAIEPFDLANQVLVRLT 1263
QY 1222 LRIGEDDHLITVMHHIISDGSWIDVLRDLNOLISA-----ALKDSKDPL-----SAL 1270
Db 1264 LVLSETEHLLVCMHHIIVSDGWSMGVFLQELAAALYNAYSQAQPSLRDATRSLLPRRGTP 1323
QY 1271 TPLTIQYSDFAKWQKDOFTE---QEKOLNWKOLKDDSPA--KIPTDFARPALLSGDAG 1325
Db 1324 APFLQYADFAWQR-QWLOGVLOSQNLWYQOQLKD-APALLSPTDRPREAVOTFAGA 1381
QY 1326 CVHVTIDGELYOSLAFRCNEHNTTGFVLLAAAFRAAHYRLTAVEDAVIGTPTIANRRPEL 1385

Db 1382 HOEFALSQKLTQGLIOLISOKQGVTLFMTLLAAAYDTLLYRYTGTEDILVSGPIANRRSEI 1441
QY 1386 EDIIGCFVNTQCMKRINDHHDTEGLINQVAKATTTAAAFENEDIPFERVVSALOPGSRDLS 1445
Db 1442 EGLIGFVNTLVNRTNLAGNPSFSEULTVRVEMAMEAYTYONLPFEMLVALQP-QRDL 1500
QY 1446 STPLAOLIFAVHSOKDLGRFKFQGLSVSP--VPSKAYTRFDMEFHLFQETDSLKGSVNFAD 1504
Db 1501 HTPLFQVMEALQN-VPLSGVELVGLRVTPLMPSERTAKFDNLNFMQNTADGLGVWEYNT 1559
QY 1505 ELFKMETVENVVRVFEILNGLQSSRTPVSIILPLTGDGIVTLEKLDLVNKHVHDYDPRESS 1564
Db 1560 DLFDASTIERMTGHFVALLGIIIANPEQIQISQLPTEVEQHOLLWEMWNTDGYDQDLC 1619
QY 1565 LADVFOQVAYSAYPDSLAVVDSSCRUTYELDRDSDILAGWLRRRSPAEFLVAVFAPRSC 1624
Db 1620 IYQLFEEQVORTPDVAVIFENQOOLYYQLNCRANOLANYLQSLGVSADVLVIGICVERSI 1679
QY 1625 ETIVAFPGVLKANLAYPLDVRSPSARVQDILSGLSGPTIVLIGHDTAPPDIEVTNVFV 1684
Db 1680 EMVVGLLGILKAGGAYVLPDPYPQORLSFILDAAALP--VLLTQOSLLESPLPHTAQVV 1737
QY 1685 RIRDALNDSNADGFEVIEHDSKPSAT-----SLAYVLTSGSTGRPKGVMIEHRIIRT 1739
Db 1738 CL-----SDSR---OLIEQHSRENPLTGSKPENLAYVIYTSGSTQPKGVQVSHNCVNF 1789
QY 1740 VTS-GCIPNYPSETRMAHMTAIFOGASYEISALLFGRTLVCVDYMTLLD-ARALKDVF 1797
Db 1790 LASMHCPCGLTSDTFCAVTTISFDIAALELILPLIVGAKVVVGVREVASDGRALLLEL- 1848
QY 1798 FREHVNAAHSVTSSSQDVP-----LRVPRRLSRTLMEFFLVVTDSTAPDADAGLYOGVQ 1853
Db 1849 --QHSS-----TTVMQATPATWQMLVASGLSTQQLGKMLCCGGEALPPQLAHOLLETAQ 1901
QY 1854 CVNGYGTENGVMSTIPI--DSTE-----SFTNG-----VPIGRALNNSGAYVVDPEQ 1900
Db 1902 VNNLYGPTETTWSICQLRNESTQLEARSATASGAALRAIPIGRPIANTQYIILDSHL 1961
QY 1901 QLVGIGVMGELVVTGDLARGYSDK--ALDENRFVH-----ITVNDQTVK 1943
Db 1962 QPVVGVAGELHTGGGLGAKVYLNRPQLTQEKFISNPFKRKRGAEQSGRGETFNSD--R 2019
QY 1944 AYRTGDRVYRIGDGLIEFFGRMDTQFKIRGNRIESAEIEAALLRDSYVDAVVLQONE 2003
Db 2020 LYKTGDLARY-LPDGNIIEYLGRIDNQVKIRGRIELGEIEAVLSQHEQVQASCVIARQ-- 2076
QY 2004 DOAPEILGEVWADHDHSENDKQOSANQVEGWQDFHESGMYSOIGEIDPSTIGDFKGWTS 2063
Db 2077 ----- 2076
QY 2064 MYDGSQIDFDEHWEHILGETTRTLHDNRSLGNVLEIGTSGMILFNLDLSRLESYVGLPSR 2123
Db 2077 -----DNPG-----DKRLVAY----- 2087
QY 2124 SAAAFVNKATESIPSLAGKAKVQVGTATDIGOVDDLLHPDLVVLVNSVIOYFPSEYLAETA 2183
Db 2088 ----- 2087
QY 2184 DTLIHLPNVORIFFGVDVRSQATNEHFLAARAHTLTKGNATKDDVRQKMAELEDMEBELV 2243
Db 2088 ----- 2087
QY 2244 EPAFTTSLKDRFPGLVEHVEHILPKMNEAVNELSAYRYAAVHVVRGSLGDELVLPEKDDW 2303
Db 2088 -----LVAHQDCKP----- 2096
QY 2304 IDFOANQLNOKSLGDLKSSDAAIMAVSKIPEITAFERQVVASLNSNIDENQLSIRSS 2363
Db 2097 -----AIS----- 2099
QY 2364 AEGDSSLYPDIFRIAGEAGFRVEYSSARQWSQNGALDAVPHCCSQGRTLVNFPTDHL 2423
Db 2100 ----- 2099

QY 2424 RGSLLTNRLQRLQNRRIAIEVRERLRSLPSYMPISNIVVLDKMPNANGKVDKRLS 2483
DB 2100 -----ELRQFLKAKLPEYWPVSAFVILESLPLTPNGKVDRR--- 2135
QY 2484 RRAKVVVKQOATAPIPTPISEVEVILCEEATEVEFGMKVDITDHFENLGGHLLATKLIS 2543
DB 2136 -----ALPK----- 2140
QY 2544 RIDORLKVRIITVKDFHPFADLASVIRQGLQOPVSDGOGQDSAHMARTETEAL 2603
DB 2141 -----DLDTTILE-----KYVAPRTPIEML 2161
QY 2604 CDEFAKVLGFO-VGTDNFFDLGGHSLMATKLAIVRIGHRLDTTVSVKDFDHPVQLAI 2662
DB 2162 ALLWAQVLKLELVGRDNFFELGGHSLATQLVSRIRTKWFKVELPLRELFSTATVFLAR 2221
QY 2663 ALDNLVQSTNEIVGREGMAEYSPQLLFTDPEEFMAEIKPQLELQELIIDIYP-STQ 2721
DB 2222 SIEQ-----FQQQDLVLVSPILPRAENAEI-----PLSYA 2252
QY 2722 MOKAFLEHHTTARPRF-VPEYIDFPSTSEPDAAGLIKACESLVNHLDIRTVFAEASGE 2780
DB 2253 QORLNFLOQFENSALYINPMALRLVGLT--NOVALEQSLYEIIRHREALRNFVTNGK 2310
QY 2781 LYQVVLSCLDLPQVI-----ETEDNINTATNEFLDEPAKPVRLGH-PLIRFT--IHK 2831
DB 2311 PSQIITOTNMKVSVGLKHLSTTEQEI--ATQQLAAQQAQIOPFDLANQALVRATLVLVS 2368
QY 2832 QTKSMRVIMRISHALYDGLSEHVVRKLMHLYNGRS-----LLP-PROFSRYMQY---- 2880
DB 2369 ETKHI-LVLCMHIIHVSQMSVFLQELAAYNACSAQTSPLAPLIQYADFALMQRQW 2427
QY 2881 -TADGRESHGFRDVIQNTP--MTILSD-----DTVVDGNDATCKALHLKIVNIPSQ 2931
DB 2428 LOGVLQSLSTWQOOLKAPALLSLPTDRPGPVQTFAGGHOEFALSVELSKNLTUKSQ 2487
QY 2932 VLURGSNIITQATVFNACALVLSRESKDSKVYFGRIVSGROGLPVEYDIOVGPCTNAVY 2991
DB 2488 E-QGCTLFMTLLAAFDT-----LLRYVTGTFEDILVGSPIANRDRSEIE--GLIGFFVNTIV 2540
QY 2992 VRAHTESS--DYNQLLHDQDQVLLSLPHETIGFSLDKNCTDWPDAITNFCSCIYHNFE 3050
DB 2541 MRTNLAGNFSFSELLGRVEMAEAYTYQNLPEML-----VEALQ-----PHRDL 2587
QY 3051 YHP--ESQPEQORVEM-----GVLTKFVNIEMDEPLYDLATAGEVEPDGAGLVTVIAK 3102
DB 2588 HTPLQVMPALQNVPLSEVELGSLVTPPLMAESITAKFDLTLL--MONTGNGLVGWVEYN 2645
QY 3103 TOLFGKRKRVEHLLVEVSKTFEGL 3125
DB 2646 TOLFDASTIKRMGSHGVTLLLEGI 2668

RESULT 13

Q9RAH4
ID Q9RAH4 PRELIMINARY; PRT: 4379 AA.
AC Q9RAH4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Nosa.
GN Nosa.
OS Nostoc sp. GSV224.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=76334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSV224;
RA Hoffmann D., Hevel J.M., Moore R.E.;
RT "Characterization of the nostopectolide biosynthetic gene cluster of
RT Nostoc sp. GSV224.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF204805; AAF15891.2; -.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; Condensation; 4.
DR Pfam; PF00550; pp-binding; 4.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
DR PROSITE; PS00455; AMP_BINDING; 4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
KW Phosphopantetheine.
SQ SEQUENCE 4379 AA; 491434 MW; 089F97BFC598E84 CRC64;

Query Match 12.2%; Score 1974; DB 2; Length 4379;
Best Local Similarity 21.5%; Pred. No. 2.1e-112;
Matches 847; Conservative 614; Mismatches 1294; Indels 1190; Gaps 130;

QY 82 IDISRFALAWKEIVNQTPALRAFATSDSGKTSQVILKDSFVFSWCMWSSSSSSPDEVDRD 141
DB 43 LNTAAVNRVWEKIIIEHPILRT--TYTIEHGKPVQV--NQCHKFNLDVANASDWSEKQLKE 100

QY 142 EAAARASGP-----RCNRFVLLEDQTKKQLVMTFSHALVDVTFQORVLSRVFAA 192
DB 101 KIFAITDRPNLEKOSILRVNLFT-----RSKKEHILLTMMHIIAGDMWFDLLSEFQL 156

QY 193 YKHEKDTHRPEPESDATTDSQSVSVSMSCEDNAVSAT-----HEFWQTHLNDLNAS 246
DB 157 YAKEIEQVSEQEIQTAPDSITENKSY-LDFVRWQSEVLSASRGEKLMQYWKOL----AG 211

QY 247 VEPHLSHLMVNP-----TTTAEHRITFPLSQKALSNSA-----ICRTALISLS 292
DB 212 ELPILNLFADKPRPVQTYQGTYSUKLDERLIQKLYLALASGTSLYQVLLAAAFVLLY 271

QY 293 RYTHSDEALFGAVTEOSLPFDKHYLAD-----GTQTVAPLRVHCOSNLRASDMVAIS- 346
DB 272 RYTNQTDILIGS-----PMRGKRGREFKEIVFVNLNVLNVSVQENATFFKEFLAQVSK 325

QY 347 -----SYDRLGHLAPFGL-----RDIRNTGNGSAACDFQVLLVTDGSH----- 387
DB 326 TVREAQKHOD-----YFSLLAPOLOPQDI-----SRSPLCQ-----VSFTQAHWCPEP 371

QY 388 VNINGINGFOQTESHSFPCNNRAL---LLHCOMESSGALLVAY-YDHNVIDSLQTLRL 443
DB 372 TENSLSHGGEQLLEIKPYLLGHOGADFOLNLMVMEAAQGVQLQLCWQYNTDLFEASTITRM 431

QY 444 LOQFGHLIKCLO-----SPDLSSMAEVNLTWEDRAE-IESWNSQPLE-VQDTLIHHEML 497
DB 432 A---GHFVTLLEGIYVNPQE--QIWQLPLLTETEQQOOLLVQWMDTGVLYLQNKCIHOLFE 486

QY 498 KAVSHSPKTKTAOAWGDWITYSELDNVSSRLAVHIKSLGIRAOQAATIPVYFEKSKWVIAS 557
DB 487 EQVDSTPNNAVAVYENQQLTYQOLNRYRANQALHYLQSLGV-GENVLVGLCVLESLEWVG 545

QY 558 MIAVLKSGNAFTLIDNPPARTAAQVVTOTRATVALTSKLHRTVOKLVGRCVVVDDELL 617
DB 546 LLGILKAGGAYVPLDPEYPAERLSEMLTDTTHVLLTQERKLVESSLPHQHQARVCLDLD-W 604

QY 618 QSVAS--DDFSSLTKSDLAYVIFTSGSGDPKGMIEHRAFSFCALKFGASIGLNSDT 675
DB 605 HSISQANQDNLSVSAENLSYVITSGTGTGPKGVAVTHQAVNRVLVNTN-YIQTAD 663

QY 676 RALQFGTHAFGACLLIMTTLINGG-----CVCIPSDDRMNSIPSEINRYNNVMMMA 728
DB 664 CVAQAANIAFDAATFEINGALLSGAKLVIITKSVLLLPPEFAVN-----LHKHKISVLF 718

QY 729 TP---SYMGTFSPEDPVGLATLVLEGEQMSSSVNAIW-----APKQLLLNGYGOSES 777
DB 719 TTALFNQLASVVPQAFSSSLRYLLFGGE---AVDPKWLVEVLKGAPO-HLLHVGYPGTEN 773

QY 778 SSIC-----FASNMTSEPNNGRAV-GAHSWVIDPNNDINRLVPTCAVCELVIESPGIA 829
DB 774 TTFSSWYLVLEDATTATTIP--IGRAIANTQIILDOOS---LQVPVPVGPGEIYLGAGLA 829


```
Db 2903 VAIGELVNSHLREIITAGEOLQITPAISQWLSKLTCTLHNHYPGSESHLATSFTLNS 2962
QY 2352 IDEWQ-----LSTIRSAEGDSSLS-----V 2372
Db 2963 VETWPLLPVGRPTANAQIYILDRFLQPVGVPGELYIAGVLLSQGFNRPELTLEKFI 3022
QY 2373 PDIF---RIAGEAGFRVSS-----ARWSON-----G 2398
Db 3023 PNPFKRSGAGEQSGRGETFNCDRLYKTGDLARYLSDGNIIEYLGRIDNQVKIRGFRIELG 3082
QY 2399 ALDAVFHCCSQGRTLNVFPDTHHLRGSDLTNRPL-----QRLQNRRIAEVRELR 2452
Db 3083 EIEAVLSQDQVASCAMATPAAGIAR-EDIPGNKRLVAYIIVPKQEKLTVSF-VRSFLKS 3140
QY 2453 LPSYMPISNVIWIDKMLPNANGKVDKRLSKRSRAKVPKQQTAAPLTFPISEVEVILCE 2512
Db 3141 KLPEYVPSAIVILDALPLTPNGKLDLR-----ALPAPDL----- 3175
QY 2513 EATEVFGMKVDITDHFPLNGHSLATKLISRIDQRLKVRITVDVDFHPFADLASVIR 2572
Db 3176 -----HSQLLDK----- 3182
QY 2573 QGLGLQOPVSDGQDRSAHMAPRTETAILCDEFKVLGF-QVGITDNFEDLGHSLMA 2631
Db 3183 -----YVAPRNPIEILSLIWAQVLKVEQVGHIDNFFELGGHSLIA 3223
QY 2632 TKLAVRIGHRLDVTTSVKVDHPVLFOLATALDNLVOSKNEIIVGGREMAEYSPFQLLF 2691
Db 3224 TQLISRVTSKVLPLSLFAAPVAIVELS---QNIQSQOQDL----- 3264
QY 2692 TEDPEEFMAEIKPOLELOEITQDIYPTQMKAFLDHTTARPPFPVFIIDFSTSEPT 2751
Db 3265 -----ELIASPLTPRAENAE-----PLSYAQTRLWFLDKLPNSAFYFPIALRLVGT 3314
QY 2752 DAAGLIKACESLVNHLDFRTVFABASGELVQVLSCLDLPQTIVTETEDNINTATNEL- 2810
Db 3315 NRTALEQSLQEIIRHEALRTNFVIVDGKPSQIIQTIQINCTLSVYDFK-HLSTTQETAS 3373
QY 2811 DEFAPKPVRLGH-PLIRFTTIKOTKSMR-VIMRISHALYDGLSLHVVYRKLHMLNGRS- 2867
Db 3374 QOLVKQPDLANEALVRATLVLSLSETHALLVCMHHIVSDGSMGVFQVOLAALNAYSQ 3433
QY 2868 -----LPL-PRQFSRYMAY-----TADGREGHGFWRDVIQNTPTIL-----SDDT 2908
Db 3434 GEPSPLTLPYADFAIQRNWLQDVLQSLSYWQHOLANAPALLSLPTDRPRPSMOT 3493
QY 2909 VVDGNDATCKALHLS-KIVNIPSOVLRGSSNIITQATVFNACALVLSRESKDVVGR 2967
Db 3494 FAGAYQOPALSKELTDKLTQLSQO--QGVTLFMTLLAAPDT-----LLRYTGTEDILVGS 3547
QY 2968 IVSGRQGLPVEYODIVGPCTNAPVVRHIES-DYNQLLHDIDQVLLSLPHETIGFSDL 3026
Db 3548 PIANDRSEIE--GLIGFLFNTVMRTNLGNPSSELLGRVREMANDAYSHONLPFEM 3605
QY 3027 -----KNCNTDWPAINFSCCITYHNFEYHPESQFQQRVEMGLVTKFVNIEDEPLY 3080
Db 3606 VEALQPERDLSHTP-----LFQVMEVLQN-----APMSQOELTGLTVSPVVKSTTKF 3655
QY 3081 DLAIAGEVDPGAGLVKVIYIAKTQLFGRKRVEHLLVEYSKTEGL 3125
Db 3656 DUTLG--MENTATGLVGIWEYNTDLFDASTIERMTGHFLTLEGI 3698
```

RESULT 14

```
Q44928
ID Q44928 PRELIMINARY; PRT: 4450 AA.
AC Q44928;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Gramicidin S synthetase 2.
GN GRS2.
```

```
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAGANO;
RX MEDLINE=95122465; PubMed=7822255;
RA Saito F., Hori K., Kanda M., Kurotsu T., Saito Y.;
RT "Entire nucleotide sequence for Bacillus brevis Nagano grs2 gene
RT encoding gramicidin S synthetase 2; a multifunctional peptide
RT synthetas.,";
RL J. Biochem. 116:357-367(1994).
DR EMBL; D29676; BAA06146.1; -
DR HSSP; P14687; IAMU.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000379; Ser_estrs_site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; Condensation; 4.
DR Pfam; PF00550; pp-binding; 4.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_2.
DR PROSITE; PS00455; AMP_BINDING; 4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_4.
SQ SEQUENCE 4450 AA; 508674 MW; F3197E77BF69316D CRC64; -
Query Match 12.0%; Score 1933; DB 2; Length 4450;
Best Local Similarity 22.0%; Pred. No. 7.7e-110;
Matches 845; Conservative 587; Mismatches 1306; Indels 1098; Gaps 137;
```

```
QY 67 KQSAIGHAVYDVPTDIDISRFALAKKEIVN---QTPALRAFAFTSDSGKTS----- 114
Db 19 QEGMLFHALLDKNAHLVQMSTATEGIVDVLLSESLNILDRYDVFRTTFLHEKIKQP 78
QY 115 -QVILKDSFVSMCWSSSSSPDEVVDEAAAA---ASGPRCNRFVLED-----MOT 163
Db 79 LOVLKERV--OLOFKDISLDEKREQAISOYKYODGETV--FOLTRPLMRVAIFQT 134
QY 164 KKC--QLVMTFSHALVDVTFQQRVLSRVAAYKHEKDHTRPETPESDATDTDSQSVSV 221
Db 135 GKVNQIMSFHILLMDGWCNFIIFNDFNLYLSKE---KKPLQLEAVQPKYKIKWL 190
QY 222 SMSCEDNAVSAATHFWQTHLDNLNASVFPHLSDHLMPNPPTT--TAHRITFP----- 271
Db 191 E---KQDKQALRYNKEHLMNYDQSV--TLPKKAAINNTTVEPAQFPAFQKVLTTQQL 245
QY 272 --LSQKALSNAICRTALSIILSRVTHSDEALFGAVTEOSLPDKHYLAD--GTYQTVP 327
Db 246 RIANSQVTLNIVFTQVIGVILQKYNSTNDVVGVSVV--SGRPFSEISGIEKMGLFNTLP 304
QY 328 LRHVQCSNLRASDVMDAISS---YDDRGLHLPFGLRDIRNTGDNGSAACDFOTVL---- 380
Db 305 LRIQTKQKQSFIELYKTVHQNVLFSQOHEY---FPLYEQNHTELQNLIDHIMVIENYP 361
QY 381 LVTDGSHVNGINGFLQO-----ITESHFMPCNRRALLHQQMESSGALL---VAYDHN 433
Db 362 LVEELQK-----NSIMQKVGFTVRYDKMFEPTNYDMTNVLPDRELSVRLDYNAAVYDID 416
QY 434 VIDSLQTTLLQOFGHLIK---CL-QSPLDLSSMAEVNLMTEVDRAE--TESNNSQPLEVQ 488
Db 417 FTRKIE-----GHMKEVALCVANNPHVL--VQDVPLTTKOEKHOHLVELHDSITEP 466
QY 489 DTLIHHEMLKAVSHSPTKTAIQAWGDGWTYSSELDNVSSRLAVHKSIGLRAQAIIIPVVF 548
Db 467 DKTIIQLFTFEQVEKTPPEHVAVVFEDEKVTYRELHERSNQLARFLREKGVK--KESIIGIMM 525
QY 549 EKSXKVIASMLAVLSGNAFTLIDPNPDPARTAAQVVTQTRATVALTSKLHRETQVKLVGR 608
```

```

Db 526 ERSVEMIVGLILKAGGAEPIDPEYKERICGYMLDSVR--LVLTQR-HLKDKFAFTKE 582
QY 609 CVVDDDELQSVASDDFSLTKSODLAYVIFSGTGDPKGMIEHRAFPSSCALKFGAS 668
Db 583 TIVIEDPSI-SHETBEIDYINSEDFLYIYISGTPKPGVMEHKNVNLHHTFEK 641
QY 669 LGINSSTRALQFTHAFGACLLIMTILNGG-CVCIPSDDDR-MNSIPSFINRYNNWM 726
Db 642 TNINFSDKVLYTTCFSFDCVYQEIFSTLLSGGQLYLIRKETQDVEQLFDLVKRENIE-V 700
QY 727 MATP-----SYMGTFSPEDEVGLATVLVGEOM---SSSVNAIWAPKLQLLNGY 772
Db 701 LSPVPAFLKIFNREFINRF-PTCVKHIIT---AGEQLVNNFEKRYLHEHNVHLNHY 756
QY 773 GQSESSI-CFASNMSTERNM---GRAVGAHGWIDPNNDINRLVPIGAVGELVIESPGI 828
Db 757 GPSETHVVTYITINPEAIPELPICKPI-SNTWIILDOEQLOQOGIVGELYISGANV 815
QY 829 ARDYIVPPPEKSPFTDIPSWPANTFPDGAKLYRTGLARVADSGSIVCLGRIDSQVK 888
Db 816 GRGLYNNQELTAEKFAD-----PFRPNERMYRTGDLARWLPDGNIEFLGRADHGVK 867
QY 889 IGRORVELGAETHLQOMPDDLTIVVEATKRSOSANSTSLIAFLIGSSYFGNRPDSDAHI 948
Db 868 IGRHRELGEIEAQLNCKGVKEAVVID---KADDRGGKYLCAVV-----MEVEV 915
QY 949 LDHDATAKINIKLEQVLRHSIPSFYICMLEPRTATGKIDRRRLRMKGKIDLDKOTQGA 1008
Db 916 NDSE---LREYLGKALPDYMIPIPSFVPLDPLTPNGKIDRKS-----PNLEGI 962
QY 1009 IVQOAPAPVPFADTAALKHSIWQSLGIDPATVNVGATFFELGGSITAIKMN-MARS 1067
Db 963 VNTNAYVYPT-NELEEKAKIWEELVGI--SOIGIQDNFFSLGSHSLAKAITLISRMKE 1019
QY 1068 VGMCLKVSNIOYHPTLAGISAVKVG-DPLSYTLIPKSTHEGPV-EQSY-----SQRLWF 1120
Db 1020 CNVDIFLRLFEAPTLOEISNYINGAKESYVIAQ-----PVPEQEYIPVSVQKRMFI 1073
QY 1121 LQOLDYGLSWLYPIYAVRMGPVNVDAARRAALAEQRHETLRTTFEDODGQVQIVHEK 1180
Db 1074 LNEFDRSGTAYNLPVGMFLDGKLYNRQLEAAVKLVERHEALRTSPHSINGEPVQRVQN 1133
QY 1181 LSEMKVIDLQSDLDPPFVNOEQTPNLSSEAGWRATLLRLGEDDHILITVMHHIIS 1240
Db 1134 VELQIAYSE---STEDQVERIIAEFMOPFALEVAPLRLVGLVKLEAERHLFIMDMHHIIS 1190
QY 1241 DGWSIDVLRDLNOLYSALKKDSKPLSALTPLPIQYSDFAKW-----OKDQFIEQKOL 1295
Db 1191 DGYSMQIMQEIADLYKE-----KELPTLGIQYKDFVWHNRLQSDVI---EKOE 1238
QY 1296 NYWKKOLKDDSPA-KIPTDFARPALLSGDAGCVHVITIDGELXOSLRAFCNEHNTTSFVYL 1354
Db 1239 AYWLNVFTEIPEVLNLPDYPRPTIQSFGKRTFTSGKQMDLYKVATETGTTLYMYL 1298
QY 1355 LAAPRAHRYLTAVEDAVIGTPIANKNRPELEDIQCFVNTQCMRNIDHDTFGTLNQ 1414
Db 1299 LAAYNVFLSKQDDIVVGTPIAGRSHADVENMLGMFVNTLAIRSLNNEDTFKDFLAN 1358
QY 1415 VRATTTAAFENEDIPPERVVVSALQPC-SROLSTPLAQILFAVHSOKDLGRKFKOGLESV 1473
Db 1359 VKOTALHAYENPDYEDTLVEKL--GIQRDLSENPLDFTWFLVQN-TDRKSFEVEQITIT 1415
QY 1474 P-VPSKAYTRFDMEFHLFOETSLKGSVNFADLFKMETVENNVVRVFFELIRNGLOSSRT 1532
Db 1416 PYVPNSRHSKFDLTLEVSEQNEILLCEYCTKLTDKTVERMAGHFLQILHAIGN---1472
QY 1533 PVSILPLTDGIVTLEKLDVL---NVKHVDYPRESSLADYFQVQVSPYSLAVWDSCLR 1589
Db 1473 PTIIISEILSEEEKHILFENDTKTTPYHMQITIQGLFEQVEKTPDQHVAVGKQDAL 1532
QY 1590 TYTELDRQSDILAGMLRRRSMAETLVAVFAPRSCETIIVAFFGLVKANLAYLPDVRSPS 1649

```

```

Db 1533 TYRELNERANQVARVLRQKQVQPDNIVGLLVERSPEMLVGMILKAGGAYLPDPPEYA 1592
QY 1650 ARVQDILSGLSGPTIVLIGHDTAPPDIEVTNVEF---VRIRDALNDNSNADGFEVIEHST 1706
Db 1593 DRISYMIQD-CGVRIML---TOOHLLSLVHDEEDCVILDEDSLYKDDSNLAPVN---1643
QY 1707 KPSATSLAYVLYTSGSTGRPKGMVIEHRVITRTVTSQICPNYPSETMAHMAITAFDQAS 1766
Db 1644 --QAGDLAYIMYTSGSTGPKGMVIEHRNIRLVKNTNYVQVREDDRIIQTAGIGFDALT 1701
QY 1767 YEIYSALLFERTLVCVDYMTTLDARALKDVFREHVNNAASHVTSSSQDVPLRPPRLSRT 1826
Db 1702 FEVFGSLLHGAELYPTKVDLLDAEKLHKL-----QANQITIMWLTSPLF--NQLSQG 1753
QY 1827 LMFEFF-----LVVTDSTAPDAL--DAQLYQGYQCYNGYGTENGVMSTIYPIDSTESFI 1879
Db 1754 TEEMFAGRLSLIVGGDALSPKHNNVKRCPNLTMNWGYGPTENTFTSTCFDID--KEYD 1811
QY 1880 NGVPIGRALNNSGAYVVDPEQLVGIVGMGELVVTGDLARGYSYDK-ALDENRFVHITVN 1938
Db 1812 DNIPIGKATISNVTYIMDRYQQLQPVGPGLCVGGDVGARGYMNQNPALTEERFVPPNFA 1871
QY 1939 DQTVKAYRTGDRVRYRIGDGLIEFFGGRMDQFKIRGNRIESATEAALLRDSVDRDAVV 1998
Db 1872 PGE-RMYRTGDLARW-LPDGTIETLGRIDQOVKIRGYRIEPEGETETLLVKHKVKVESVIM 1929
QY 1999 LQONEDQAPILGVFVADHD-----HSENDK-----2024
Db 1930 VVEDNNGORALCAYVPEEBEVTVSELREYIAKELPVYWPAYFVQIQEOMPLTQNGKVNRS 1989
QY 2025 -----GOSANQVE-----GWQDHF-----2038
Db 1990 ALPKPDGEFTATYEVAPSSDIEMKLAIEIWHNVLGVNKIGVLDNFFELGGHSLRAMTMS 2049
QY 2039 -----ESGMYSIDGIDPST-----IG 2055
Db 2050 QVHKFEFDVPLKVLRETPTISALAQYIADGEGMYLAIQPVPDQYYPVSSAQKRWYL 2109
QY 2056 SDFKGWTSYDGSQIDFDE-----MHEWL-----GETTRTLHDNR 2090
Db 2110 YEPEGAGITYNVPNVMFIEGKLDYQRFYAKSLINRHEALRTSFYSLNCEPQVRVHQV 2169
QY 2091 SL-----GNVLEIG--TGSMILFNLSR-----2112
Db 2170 ELQIAYSEAKEDIQVIESFVQPFDEIAPALRVGLVKLASDRHLFLMDMHIIISDQVS 2229
QY 2113 -----LESYVGLEPS-----RSAAA-----FVKATESIPSL--2139
Db 2230 MQIITKEIADLYKGKELAEHLIOYKDFAVWQNEFWFQSAALEKOKTYWLTFAEDIPVLNL 2289
QY 2140 -----AGKAK 2144
Db 2290 STDYPRPTIQSPEGDIWTF-SAGKOLABELKRLATETGTTLYMLLLAAYNVLLHKYSQEE 2349
QY 2145 VQVGTA-----TDICQVDDLHPD-----LVVLNSVITQYFSPSEYLA 2180
Db 2350 IVVGTPIAGRSHADVENIVGMFVNTLAKNTPIAVRTPEHFLLEVKQNALEAFENODYPE 2409
QY 2181 E-----IADTLHLPNV-ORIFPG-----DVR 2201
Db 2410 ENLIEKLOVRRLSRNPLFDTWFSLSNIDEQVEIGLSFSFYEMOYIAKEDIISFDIL 2469
QY 2202 SOATNEHF-----LAARAIHTLGKNAKDDVRQKMAEL*---EDMEEEL 2241
Db 2470 EKODDIQYFNVCTNLFKKTETTERLATHEFMHILQEIWINPEI--KLCEINMLSEEQQRV 2527
QY 2242 LVE-----PAFFTSLKDRFPGLVEHVEILLKPNMAVNELSARY-----2280
Db 2528-LYDFNNGTDATYATNKFHELFEQVEKTPDHIADVIDEREKLSQELNAKANQARVLRQK 2587
QY 2281 -----AAYVHVIRG 2288
Db 2588 GVQPNMVGIMVDRSLDMIVGLVLCAGAYVPIQIDYPOERISTYMMEDSGALLITQO 2647

```


QY 161 -----MOTKKCOLWTFESHALVDTFOQVRLSRVFAAYKKHEKDH-----RPEPPES 207
D 130 LLRFTALRSADRVRVMTNHHVLDGWSMPVLLRELMAALYAAEGDPSALPRVPRDY 189
QY 208 S--DATDTSQSVSVMSCEDNAVSATHFWQTHLNDLNASVPFPHLSHLMVFN--PTTT 263
D 190 AWDARD-----AARDWRRSLSGLEAT-----LLAPDGPAST 226
QY 264 AEHRITPLSQ-----KALNSAICRTALSILLSRYTHSDLEALFGAVTE--Q 308
D 227 APSQVTFVDEYSGALSAMARGCVTNTVVGANALALAAQATGRDDVVFGATVSGRPP 286
QY 309 SLFPDKHYLAGTYQTVAPLRVHCQSNLRASDYMDAISYDDR-LGLAPFLGLDIRNTG 367
D 287 ELPGVESMI--GLFINTPLVRARLDQAEPLGLDFRRLQNEQARLLDHPW-GLADIQHW 343
QY 368 DNGSACDFQTVLLVDTGSHVNGINGFLQIITESHFMCNCRALLHCHQMESSGALLV 427
D 344 GHGEL---FDTAMVFN-----YPVEGDLTAPADPDL-----RVASADIKGG 384
QY 428 AYVDHNVDSLOQTRL-----LOQFG-HLIKLOSPLDLSSMAEVLNMT 471
D 385 THEAVNVATMGAELSFRVDYRDLDEAYARDFGRMLRVLET-----LISD 433
QY 472 YDR--ABIES-----WNSQPLEVQDTLIIHEMLKAYSHSPTKTAIQAWGDWT 517
D 434 PDRPVAHLDPDAVRERVLVEVNGAPTQLPGTPLHELISEQARLTDPDAVAVCDGTSLT 493
QY 518 YSELDNVSSRLAVHKSGLRAQOALIPVYFEKSKWVIAWMLAVKSGNAFTLIDNDPP 577
D 494 YAEIDGRANGLARHLERLUGAED-FVAIALHSLDAVMTSLAVLTGAAYLPIDPDPA 552
QY 578 ARTAQVVTQTRATVALTSKL-----HKTQVKLVGRVVDDELQSVASDDFSLSLK 631
D 553 ERITYMLDDARPALTLREPVPAAAYGHRPT-----DD--VTDAERTPWSAL-- 597
QY 632 SODLAVIFTSSTGDPKGMIEHRAEFSKALK-----FGASLGINSSTRALQGTGTAFGA 687
D 598 --HAAYMYTSGTGRPKGVVIEHHALATVHLHARTYNTAMTG-----TVLHSPFLAFDL 650
QY 688 CLEIMTTLINGCVCIPSDDDRMNSIPSPINRYNVMNMATPSYM-----GTFSPED 740
D 651 TITALTPTLGTAGTVHLSLEE-AEVQPSLIK-----ATPSHPLTLNLAATASPSH 701
QY 741 VPGLATLVVGEQSSSVNAIWA---PKQLLNGYQSESS-SICFASNMTSE--PNNMG 794
D 702 -----TLILGEALHTDQTDWRTQHPGAQIINAYGPTSTVNITDHRLDGTEEGVPVIG 756
QY 795 RA-VGAHSWIDPNDINRLVPI--GAVGELVIESPGIARDYIVPPPEKSPFTDIPSWY 851
D 757 RPFANTQVYVLD-----SALRPVAPGTGELYLAGELQARGYLGRPALTAERETAN----- 807
QY 852 PANTFPDGAKYRTGDLARYASDGSIVCLGRIDSQVKIRGORVELGAIETHLRQOMPDDL 911
D 808 PHSSVP-GARMYTGDLAHNHHGHTYDGRADHQIKLGRHREPEIETTLTAQ----- 861
QY 912 TIVVEAT--KRSOSANSTSLIAFLIGSS-YFNGRPSDAHLDHATKAINIKLEQVLP 968
D 862 TGITOATVOLREDPQDQRLVAVLVSTGVDENTVRDA-----LVLSALPDY 907
QY 969 SIPSFYICMLELPTATGKIDRRRLIMGKIDLDKOTQGAIVQOAPAPVPF-ADTAAK- 1026
D 908 MVSALVTLDAPLTPNGKLDRTAL-----PAPYASASTGRT 945
QY 1027 -----LHSINWQSLGIDPATVNVGATFFELGGNSITAIRMVNMAR-SVGMCLKVSN 1077
D 946 PRTPREILLTXSEILSVN--TVGIDDSFFDLGGHSLATRLVSRIRTLGRELIRQL 1003
QY 1078 YQHPTLIAGISAVVKGDPLSYTLIPKSTHEGPEVQSYSGRLMFLDOLDVGSWLYLIPYAV 1137
D 1004 FETPTVAGLSRALDTSGLTALTARPRPRIPRLPSVAQORLWFLHOLEGPTATYNIPTTL 1063

QY 1138 RMRGPVNVADLRALAALEQRHETLRTTFEDQDQGVQIV--HEKLSSEMKVIDLCSGL 1195
D 1064 RLGTGLDLDALOSAFNDLLARHESLRTTYTEDDQGAQIVLPVEAVATPFAVVDVAAEDV 1123
QY 1196 DPEVLNQEQTTFPNLSSEAGWRATLRLGEDDHLIIVMHHITSDGWSIDVLRDLNQL 1255
D 1124 --AERVAEAAAHAFDLGAETIPVRLFRVSEREHVLLLLVHHITASDWSRGPLAQDLTAA 1181
QY 1256 YSAALKDKSDPLSALTPLIOYSDFAKWOKDOFTEQ-----KOLNYKKOLKO--SSP 1307
D 1182 YTARCADD--APAWQPLPVQYADYALWQOEIIGDDTDPDTLAGRQAYWKQOGLAGLEP 1238
QY 1308 AKIPTDFARPALLSGDAGCVHVTIDGELYQSLRAFCEHNTTSEVVLAAAFRAAHYRLTA 1367
D 1239 LDLPDTPRPAATADHTGDRVTFTVPADLTRLTELARETNTTAFMWIQAALALLTRHGA 1298
QY 1368 VEDAVIGTPTANRRPELEDIIGCFVNTQCMRINIDHHDFTGFLINOVKATTAAFNED 1427
D 1299 GEDIPGTVPAGRTDDATDHLIGLFFVNTLVLRDTSGNPTFRDILLTRVRDITLAYTHQD 1358
QY 1428 IPEVVVSALQPGSRDLSSTPLAQILFAVHSOKDLGRFKOGLESVPVPSKAYT-RFDM 1486
D 1359 LPERLVEALNP-TRSLTHPLFQTMLTHTNOGTKDRFAGLAAEVVASESVSARDLS 1417
QY 1487 FHLQETDS-----LKGSYNFADELFKMETENVVRFFELIRNGLOSSRTVPSILPLT 1540
D 1418 FALAHEFGADHSCAGSGGVTYSTALFDRATVRDLADRLVRLLLAAVAHPGRSVGOLEIM 1477
QY 1541 DGIVTLEKLDVANKVHDYPRESSLADVFQTVSAVPSDLSLAVVDSSCRLYTETLDRSDI 1600
D 1478 DAERLVLQEWNDTAAE-PPAASVTGLFERQARRSPGATLEFGEVRLSTAEULANRANR 1536
QY 1601 LAGWLRRRSPAPETLVAVFAPRSCETIVAFPGVLKANLAYLPDLDVRSRPSARVQDILSGLS 1660
D 1537 LARHLVARGAARCFVAVALPRSAELVVTLAVLKSCAAVLPIDPHYPADRVEVMLAD-A 1595
QY 1661 GPTIVILIGHTDAPDIEVTNVEFVRIRDALNSADGFEVIEHDSKPSATSILAYLYTS 1720
D 1596 GPALTV---TEP---VAEAGLSGYGDA--DLGADELRGPVHGAHP-----AYTIYS 1639
QY 1721 GSTGRPKGVMIHRVIRTVTSCIPNY-----PSETRMAHMTAFDGCASVEY 1770
D 1640 GSTGRPKGVV-----VPRGALDNFLADMGRRTPGSGDRLLAVTTVGFDIAGLEIF 1690
QY 1771 SALLFGRTLVCVDYMTTLDAARLKDVFREHVNAAHVT-----SSSQDV 1815
D 1691 LPLHGAVLVLADEETARDPHALL-----HRVSASGITVMVQATPSLWQGVAAVAGDELA 1744
QY 1816 PLRV-----PRLSRTLMFFFLVVTDTAPDALDAQGLYQGVQCYNGYGPTEGVMS 1868
D 1745 GVRVLVGGALPSELARAL-----TD-----RARSVTNLYGPTTEATIWAT 1784
QY 1869 IYPIDSTESFINGVPIGRALNNSGAYVVDPEOQLVGIGVMGELVVTGDLGARGYDK-AL 1927
D 1785 --AADVAES---GPVIGRPLANTSAVLDLSALRPVGVPGELYLAGELAQYHLRPA 1839
QY 1928 DENRFVHITVNDQTVKAYRTGDRVRYRIGDGLIEFFGRMDTQPKIRGNRTIESAIEAALL 1987
D 1840 TSEFTADPVPAGTARMYRTGDLV-CRRRDGALRYLSRVDDQVKLRGRIELGEIEAELS 1898
QY 1988 RUSSVPRDAVVLQONEDQAEIILGFVVADHDHSENDKQOSANQVEGHDFESGMYSIDC 2047
D 1899 RHPVAESA-VA-----RHPVAESA-VA----- 1908
QY 2048 EIDPSTIGSDFKGTSMYDGSQIDFDMHEWLGETTTRTLHDNRSLGNVLEIGTSGMILF 2107
D 1909 -----TVREDRC----- 1916
QY 2108 NLDRLSLEYVGLPEPSRAAFAVFNKATESIPSLAGAKAVQVGTATDIGQVDLHPLDLVLN 2167
D 1917 --DRRLVGIV----- 1924
QY 2168 SVIOYFSPSEYLAETLHLPNVQIRIFGDIVRSQATNEHFLAARAIHTLGNKATKDDV 2227

Db 1925 ----- 1924
QY 2228 RQMAELEDMEBELIVERAFTSLKDRPGLVEHVEILPKNMEAVNELSAYRYAAVVHVR 2287
Db 1925 -----VPGPE----- 1930
QY 2288 GSLGDELVPVEKDDWIDFOANLNOKSLGDLKSSDAALMAVSKIPFEITAFERQVVAS 2347
Db 1931 ----- 1930
QY 2348 LNSNIDENQOLSTIRSAEGDSSLVDPDIFRIAGEAGFRVEVSSARQWSONGALDAVFHHC 2407
Db 1931 -----GPAGTGPVPSAR----- 1943
QY 2408 CSQGRTLNFPDHHLRGSDLLTNRLQRLQNRRTAIEVRERLSLLPSYMTIPSNIVVLD 2467
Db 1944 -----ELRERLGRLPFEFVPTAFVVD 1966
QY 2468 KMLNANGKVDKELSRRAKVVVKOQTAAPLTFPISEVEVILCEEATEVFGMKVDITDH 2527
Db 1967 ALPLTPNRKLDK-----ALPAPEY----- 1986
QY 2528 FENLGHSLLATKLISRIDQRLKVRITVVDVDPVADLASVIRQGLQOPVSDGQ 2587
Db 1987 -----DGEV 1991
QY 2588 DRSAHMAPTETEAILCDEFKVLGF-QVGIITDNFFDLGGHSLMATKLAVRIGHRLDTTV 2646
Db 1992 GR-----GPRDPEELICALFAEVGLVGARVIGDGGFDLGGHSLLATRLVSRITLGVEL 2047
QY 2647 SVKDVDFHVLQALADNLVQSKTNEIVGGREMAEYSPFOLLETPPEEFMASEIKPQ 2706
Db 2048 SVROFFETPTIAGLSGALD-----RAAGARAP-----LAARPRE 2082
QY 2707 LELQLIQDIYSTOMOKAFLEDHTARTPRPVFPYIDPPTSEPDAAGLIKAC-ESLVN 2765
Db 2083 R-----VPLSPAQQRLWLFHQFEGPSATYNNPTALRLSGPVDRAALERAIGDVLVR 2133
QY 2766 HLDIFRTVFAEASGELYOVVWSC-----LD-LPIQVITETEDNINTATNEFLDEFKPEV 2818
Db 2134 H-ESLRTVFAADGGSGWVLPADRAVGRDLVDVDTAGEVARGVGEAAHAFDLTADIPF 2192
QY 2819 RLGHPLIRFTTIKQKSMRVIMRISHALYDGLSLEHVVRKKHMLYNGR--SLLP-----P 2871
Db 2193 -----LARLFRVSDTEHV-LLLIHHIAGDGWSMAPLARDLTAAYAACAGAAPDWEPLP 2246
QY 2872 HOFSRIMQYTAD--GRESGH-----GFWRDVIQNTPTMTILSDDTVVVDGNDATCKALH 2921
Db 2247 VOYADYALWQREVLDGSDPDVAAAROLAYWKDALAGLPEQL---ELPTDRPRPATAGYT 2303
QY 2922 LSKVINIPSOVLRGSSNIITQAT-----VFNAACALVLSRESKDVVFGRIVSGROGL 2975
Db 2304 GDRIFTVPTALHARLTALARSTHSLFMVQAAATLTLRLGAGEDIPVGTAVAGRND 2363
QY 2976 PVEYQDVGPCNAVVPVRAHIESS-DYNOLLHDIQOYLLSLPHETIGFSDIL-----KR 3028
Db 2364 ATE--DLVGFFVTLLVRLTGTGNTFRELLGRVREDLAAVAHQDVPERLVEALNPAR 2421
QY 3029 NCTDWPPEATNFSCCITYHNFYHPESQEQVEMGVLTKFVNEMDEP----- 3078
Db 2422 SLAHP-----LYQVMITFNN-----TAGAGDRTPASPDPDVSQAGALL 2463
QY 3079 -----LYDLAIA-GE-VEPDG--AGLKVTVIATKTLFGKRKRVEHLEEV 3118
Db 2464 NATRMTAGTVAKFDLALTFGRHDPGGPGAGRSLETRTELFDRETAESVISRL 2519

Search completed: May 30, 2003, 12:48:34

Job time : 245 secs

THIS PAGE BLANK (USPTO)